ASSESSING THE DETERMINANTS OF CARE-SEEKING FOR CHILDHOOD ILLNESS IN RURAL PUNE DISTRICT, MAHARASHTRA STATE, INDIA

by
Andrew D. Marsh, MPH

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
September 2018

Abstract

An estimated 1.2 million children under five years of age die each year in India, with pneumonia and diarrhea among the leading causes. Interventions exist to reduce mortality and morbidity from these causes, though approaches to measure intervention coverage typically rely on respondent self-report and may be subject to bias. Technology-based approaches, such as those using Global Positioning System (GPS) data, provide an alternative while potentially reducing these biases. Recently, the Improving Coverage Measurement (ICM) study compared maternally-reported care-seeking for childhood illness with care-seeking assessed through a GPS-based approach in rural Pune district, Maharashtra, India. We analyzed data collected through the ICM study to:

1. Evaluate the association between care-seeking for childhood cough, fever, or diarrhea and child, maternal, and household factors
2. Evaluate the feasibility of a smartphone-based approach for tracking participant movement and explore factors associated with data completeness and participant compliance with phone-related protocols
3. Evaluate the positive predictive value of a GPS-based method to detect health facility visits and explore the factors associated with method's performance

Care-seeking for childhood illness was high overall and most care was provided through the private sector. Characteristics of the illness, especially perceived severity, were the primary contributors to the decision to seek care. Maternal employment was associated with decreased care-seeking for illnesses perceived as non-severe though not associated with care-seeking for more severe illness. Smartphone-based movement tracking through a location-aware application resulted in both high data completeness and participant compliance. Data completeness increased among participants complying with phone-related protocols and residing in rural villages. Compliance increased during the second half of the study and was highest among participants of higher socioeconomic status. Overall performance of the GPS-based method to detect health facility visits was low with most detected visits subsequently classified as false positives. The probability of detecting false positives increased among participants who spent more time in urban centers and at facilities located nearer to participants' homes. Alternative approaches to monitor health behavior may be preferable while GPS-based approaches are refined. As their performance improves, smartphone-based approaches may provide a platform to integrate behavioral observations with other smartphone applications.

Thesis Advisor:

Robert Black, MD MPH
Professor, Department of International Health

Thesis Readers:

Joanne Katz, ScD
Professor, Department of International Health

Cynthia Minkovitz, MD
Professor, Department of Population, Family, and Reproductive Health

William Moss, MD MPH
Professor, Department of Epidemiology

Alternate Thesis Readers:

Larry Chang, MD MPH
Professor, School of Medicine

Shivam Gupta, MD MPH PhD
Associate Scientist, Department of International Health

Acknowledgments

This thesis represents the culmination of several years' work. While it holds special significance for my personal and professional development, this would not have been possible without the support of many, many people.

First, I would like to thank the study participants and study staff at the Vadu Health and Demographic Surveillance System (HDSS). This research would have been impossible without your efforts. I made my first trip to the Vadu HDSS in November 2014 and can still recall meeting Sandeep by the road in Viman Nagar where we joined the vehicle that would bring us to the field site. The field staff welcomed me as an honorary member of the office, even to the point of providing me an engraved name placard to reserve my desk space. I would like to thank Shirish, Pardeshi Sir, and the many others who helped me to feel at home during my visits.

This study was particularly ambitious and included components that were new to many of us. Even so, the team remained well-organized and went beyond expectations to implement this study. This would not have been possible without the dedication and leadership of Ms. Pallavi Lele and Ms. Uddhavi Chavan and to the hard work of all of the field research assistants. In addition to your contributions to making this project a success, I also want to recognize the hospitality that you both showed at various times during my stay. I am glad to have worked alongside you.

Separately, I would like to recognize some of the senior researchers among our team for the many ways in which they supported me through this journey. I would like to thank Dr. Sanjay Juvekar, Officer-in-Charge of the Vadu HDSS, for the warm welcome he showed me when I arrived and for his continued hospitality throughout my stay. Your door was always open to me and I appreciate the many times that you sat with me brainstorm solutions to the various things we had encountered in the field. Next, I would like to thank Dr. Siddhi Hirve, former Officer-in-Charge of the Vadu HDSS but also a classmate of my father's when the two were pursuing the Masters in Public Health at Harvard. Siddhi, you have a talent for asking difficult questions and we have all benefitted tremendously as a result. In one of our first encounters you warned me that you did not plan to take it easy on me with regard to my PhD analyses. I'm inclined to say you have kept your word and the result is all the better for it. Your acknowledgment of Tofu inspired me to do the same for our family cat. Chuck, you may never read this but you have been a blessing for the entire family. Keep up the good work.

In addition to the contributions of Drs. Juvekar and Hirve, this study benefitted immensely from the expertise of Drs. Harish Nair and Harry Campbell of the University of Edinburgh. While my home campus was in Baltimore, your support throughout this process made me feel as though I my PhD was a joint collaboration with the University of Edinburgh. I have come to discover just how much the two of you travel for work, which makes it all the more incredible that you attended nearly all of our weekly teleconferences and even volunteered to keep the minutes. Your continued investment in the study truly left an impression on me.

The Johns Hopkins Bloomberg School of Public Health has been my home base these last several years. There are many faculty here and elsewhere in the university who have contributed to my development. I would like to recognize the following individuals for their support along the way: Cynthia Minkovitz, Bill Moss, Joanne Katz, Larry Chang, Shivam Gupta, Will Checkley, Alain

Labrique, Krishna Rao, Neff Walker, Marie Diener-West, Luke Mullany, Danny Feiken, and Larry Moulton. I would also like to thank Karla McCarthy and Cristina Salazar for all their support.

I arrived at the School of Public Health by way of the Institute for International Programs (IIP). I owe a huge debt to Jennifer Bryce, who first took a chance on me by bringing me on to help with the costing for the Real-time Mortality Monitoring project. Through that project I became affiliated with the costing for the Catalytic Initiative to Save a Million Lives, eventually leading me to my current project and the basis for this thesis research. I have been honored to work beside so many brilliant researchers. First and foremost, I would like to thank Jennifer for so expertly filling the role of boss, mentor, and friend, sometimes all at once. At the time of writing, the answer to your perennial question is "not yet," but please keep asking. Next, I would like to thank my advisor, Bob Black. I faced several challenges through this process and was always able to count on your support. I have been amazed by your combination of efficiency and insight. More than once I have sent you an e-mail on my way to work and reached my desk only to see a reply already waiting in my inbox. What's more is that your replies always get right to the point of whatever is being discussed. It has been such a privilege to move through this process under your guidance. I would also like to thank the many faculty working with IIP, including Melinda Munos, Neff Walker, Agbessi Amouzou, and Rebecca Heidkamp, among many others.

Finally, I would like to thank my parents, David and Christine, my two brothers, Eli and Dan, and their families. As the youngest child, I can definitively say that you've supported me from my first breath. You have been understanding through countless missed birthday parties, family reunions, and other events when I had to put life on hold to take care of work. You have been a source of

encouragement and motivation whenever those were lacking. I look forward to life's next chapter and seeing what it has in store for us.

Table of Contents

Support

This work was supported in part by a prime award by the Bill and Melinda Gates Foundation (grant no. OPP1084442) made to the Johns Hopkins University of which University of Edinburgh, Edinburgh, UK and King Edward Memorial Hospital Research Centre, Pune, India were sub-grantees.

List of Tables

List of Figures

Abbreviations

ABM	Andersen Behavioral Model of Health Services Use
ALRI	Acute lower respiratory infection
ANM	Auxiliary nurse midwife
AOR	Adjusted odds ratio
API	Application programming interface
ARI	Acute respiratory infection
ASHA	Accredited Social Health Activist
AUC	Area under the receiver-operator curve
BMGF	Bill and Melinda Gates Foundation
CI	Confidence interval
DHS	Demographic and Health Surveys
DTP	Diphtheria, tetanus and pertussis
GIS	Geographic information system
GPS	Global positioning system
HF	Health facility
ICDS	Integrated Child Development Services
ICM	Improving Coverage Measurement in Maternal, Newborn, and Child Health
IIP-JHU	Institute for International Programs at Johns Hopkins University
IMEI	International mobile equipment identity
IQR	Interquartile range
IRR	Incidence rate ratio
KDE	Kernel density estimation
KEM Hospital	King Edward Memorial Hospital
Km	Kilometer
LMIC	Low- and middle-income country
MICS	Multiple Indicator Cluster Surveys
NFHS	National Family and Health Survey
NGO	Non-governmental organization
OR	Odds ratio
OS	Operating system
PHC	Primary health center
PHFI	Public Health Foundation of India
PR	Prompted recall
REF	Reference group
RR	Relative risk
SD	Standard deviation
SES	Socioeconomic status
SIM	Subscriber identification module
SSI	Semi-structured interview

ST-DBSCAN	Spatial-Temporal Density-Based Spatial Clustering of Applications with Noise
UNICEF	United Nations Children's Fund
UoE	University of Edinburgh
Vadu HDSS	Vadu Health and Demographic Surveillance System

Chapter 1: Introduction

Public Health Importance of Care Seeking and its Determinants

An estimated 1.2 million children under five years of age die each year in India, with pneumonia and diarrhea among the leading causes of mortality [1]. Effective treatment and prevention options exist to reduce mortality and morbidity from these causes and have been estimated to be able to reduce global pneumonia and diarrhea-specific mortality by 65 and 88 percent respectively if increased to universal coverage [2]. The Pathway to Survival framework identifies care seeking as a key step between onset of illness and the restoration of health, especially in contexts where the necessary treatment cannot be provided within the home [3].

While several models explain the determinants of care seeking, a model developed by Andersen is among the first and the most widely used [4]. The first version of the Andersen Behavioral Model of Health Services Use (ABM) hypothesized that certain predisposing, enabling, and need characteristics could explain an individuals' variability in health services use. Predisposing characteristics predict an individual's propensity to seek care prior to the onset of illness, enabling resources make care accessible to an individual, and need factors describe the level of illness an individual is experiencing. Through a series of revisions, the model was expanded to include factors relating to the health system, consumer satisfaction, health outcomes, and personal health behavior [5-7].

While most of the ABM's applications have been in developed countries, Kroeger provided a framework for operationalizing the model in the context of health service research in low- and middle-income countries (LMICs) [8]. Figure 1.1 presents the adapted version of the ABM put forward by Kroeger. For each independent variable, Kroeger draws from the literature to describe the association with care seeking. These findings are consistent with those of more recent studies to address the determinants of care-seeking for childhood illness in LMICs. Bennett and colleagues conducted a meta-analysis of 258 nationally-representative household surveys conducted in LMIC settings, identifying the following common determinants of care-seeking for childhood illness: child age, child sex, maternal education, urban residence, household socioeconomic status, and distance to the nearest health facility [9]. Similarly, Geldsetzer and colleagues conducted a systematic review of studies on the recognition and care-seeking for childhood illness in developing countries, reporting the following determinants as a secondary objective of the analysis: geography (urban residence, distance to the nearest facility, and relative proximity of formal health providers relative to traditional healers), illness severity (caregiver perceived severity and clinically evaluated severity), socioeconomic status and cost of health care, and child sex [10]. For a more detailed discussion of care-seeking determinants within the study area of this thesis research, see Section 3.1.2 ("Study Population").

The Relationship between Distance and Care Seeking

Geographic access to care represents an important determinant of care seeking practices. Kroeger includes access as a characteristic of the service, while the original version of the ABM includes access to care as an enabling characteristic within the category of community resources [4,8]. As with other enabling resources such as household wealth, geographic access to care facilitates the care seeking process. There are several approaches to measure geographic accessibility, including the number of facilities within a given spatial unit, the number of facilities within a given distance from a specific point, and travel impedance to the nearest facility (measured as either distance or time) [11,12].

The simplest calculation for travel impedance is the straight-line distance between two points (Euclidean distance). This approach does not take into account geographic features that hinder travel (e.g. rivers, mountains) or features that may enable travel (e.g. road networks). More sophisticated approaches to assessing travel impedance account for these features, such as network-based methods that calculate the shortest path along a road network. The added complexity of these approaches provides more accurate measurements of access to care, though the relative improvement depends on the study context. Nesbitt and colleagues compared six methods of varying complexity to assess travel impedance, with Euclidean distance being the simplest [13]. The authors found high correlation among all tested methods and suggest that Euclidean distance may be an appropriate measurement of geographical access in certain settings, such as those where the geographic features described above (e.g., roads, rivers, road networks) are evenly distributed. Alternative measures, such as network-based distance, may be more suitable where the study environment does not support Euclidean distance as a measure of access. Nesbsitt and colleagues cited three LMIC that found Euclidean distance to perform poorly when compared with more sophisticated approaches, two in rural East African settings with substantial geographic variability and a third in a highly urbanized South American city [13-16]. These findings support the context-specific nature of accurately measuring access to care. Network-based methods generally provide a more accurate measurement, though the additional inputs required (e.g. road network, GIS software) may offset that incremental benefit. Selecting the best approach will depend on the research question, study setting, and the availability of auxiliary inputs required by more sophisticated approaches.

A review of studies evaluating the association between travel impedance and utilization of child health generally supports the inverse relationship between distance and utilization. (Table 1.1) These studies assess a variety of outcome measurements. Four studies assessed association between distance and the decision to seek care, all of which found significant associations between increased distance and decreased likelihood of seeking care for fever, acute respiratory infection, or malaria/fever [9,17-19]. In one of the studies, the effect of distance on returning for care for persistent illness was greater than the effect of distance on initially seeking care [17]. Three studies evaluated the association with the delay in seeking care after illness onset and quality of care received, measured as provision of the appropriate medicine [20-22]. Two studies assessed the association with the level at which care was sought, with one study reporting that care was more likely to be sought from a traditional healer as the distance to the nearest health center increased [23,24]. Of the remaining two studies, one reported a reduced incidence rate of pediatric clinic attendance for all causes among children living farther from the health facility, while the other observed no significant relationship between distance and the time at which a child received his/her third dose of DTP/DTP-pentavalent vaccine [25,26].

These studies measured travel impedance as both distance and travel time. The studies took several approaches to the calculation of these travel impedance measures, though the information was primarily assessed through an interview of either study participants or clinic staff [19,20,22-24]. Two studies calculated distance from the nearest health facility using participant-level GPS coordinates and one used village-level GPS coordinates [9,25,26]. Participant-level GPS coordinates provide a more precise method to both classify distance and identify the nearest provider when compared to village-level coordinates, which were found to incorrectly identify the closest facility to a household 13% of the time in a recent study [13]. For the remaining three studies the method of calculating travel impedance was not specified [17,18,21].

The Use of GPS in Health Research Studies

Common approaches to assess care seeking and other health behaviors include surveys, observations, and diaries. However, these approaches are subject to various biases that decrease their validity and increase the uncertainty about their resulting measurements. Surveys allow for retrospective assessment of behavior, usually within a specific recall period, though they are subject to the validity of participant responses due to poor recall or non-completion [27]. Diaries and direct observation both allow for prospective assessment of participant behavior, though diaries face issues of missing or incomplete data while direct observation is often impractical due to issues of cost and privacy. Tracking human mobility using portable GPS devices presents an alternative method of assessing health behaviors [28].

Technology-based approaches, such as those based on Global Positioning System (GPS) data, provide a minimally invasive approach to continuously monitor participant behavior while minimizing the biases inherent to more traditional observational methods [29]. GPS receivers communicate with orbiting satellites to determine their location with a precision of up to a few meters, depending on the number of visible satellites and their relative placement in the sky [30]. GPS has traditionally been measured using dedicated devices, though it is suggested that the single function of these devices makes them less attractive to study participants and results in decreased compliance with study protocol [31].

Recently GPS receivers have become commonplace in many commercially available smartphones and have been shown to produce comparable data to traditional GPS devices across various settings [32-37]. Traditional GPS signal acquisition methods operate independently of the cellular network, enabling location estimation even when no cellular signal is available. Where coverage exists, smartphones use cellular network data to supplement traditional methods and determine a location fix more quickly than GPS receivers [34]. GPS-enabled smartphones have not only been applied extensively in the field of transportation to model travel mode [38] and route choice [39], but also used to characterize participant life space [40] and social engagement [41] and to describe the context of participant activities, such as environmental exposure to air pollution [42] and hospitalization [43].

Two notable studies have applied a smartphone-based approach to the collection of participant movement data over an extended period of time. The Human Mobility Project provided its application for free online download, enrolling 270 participants from 13 countries and collecting an average of 6.3 days of location data each [44]. Glasgow and colleagues provided a smartphone application to 42 participants, collecting an average of 122.7 days of observation each [42]. A shared feature of these two studies is their use of participants' currently owned smartphones for their applications, greatly reducing study implementation cost. This may be an especially important feature in contexts where there are significant costs associated with the subscription to a mobile data network. In such a setting, integrating an application within a user's existing mobile phone and mobile data plan may substantially reduce overall study cost. While predominantly taking place within the context of developed countries, these studies demonstrate the feasibility of applying a smartphone-based approach to the large-scale collection of participant movement data both in terms of the number of participants included and the total duration of study follow-up.

Several studies identified issues related to variable quality among smartphone models [33,34,37], highlighting the importance of device selection and testing prior to study implementation. Additionally, smartphones permit real-time data transfer and communication with research participants, though device customizability, including the installation of additional applications and the configuration of device settings, introduces the potential for non-standard device performance [28]. It is possible to restrict these functions, but doing so may reduce attractiveness of the device to participants [45].

While GPS devices present a promising approach to study mobility and human behavior, there are limitations to their use. A review of 24 studies applying GPS to the study of physical activity found data loss between 3% and 92%, highlighting the importance of device selection (e.g. with regard to battery life, size, portability, and weight) and efforts to maximize participant compliance [46]. Additionally, GPS performance is influenced by features of both the natural and built environment and may deteriorate in areas where large buildings obstruct satellite visibility [29]. While smartphones have been shown to provide location data of comparable quality to traditional devices, there is substantial variability between models, resulting in some concern about the quality of these devices relative to devices such as data loggers and GPS transmitters. This reduction in data quality has been observed in at least one context, possibly explained by suboptimal hardware configuration (e.g. GPS antennae located in a part of the phone that is obstructed by other components) and different storage behaviors (e.g. phones are typically stored in pockets or handbags, where signal may be worse than if worn or carried) [47]. Maas and colleagues reviewed the advantages and disadvantages GPS-enabled smartphones and four other device types for monitoring participant behavior [28]. The review highlights the potential for GPS-enabled smartphones to transmit location data in real-time while also noting that smartphones are limited by relatively high cost and the potential for users to manipulate settings [28]. Traditional GPS data loggers cost less than smartphones and are less susceptible to participant manipulation of settings but require that data be periodically downloaded directly from the device. Previous applications of GPS-enabled smartphones have predominantly occurred in developed countries and little is known about whether a similar approach would be feasible in a less-developed context.

Analytical Approaches for Participant GPS Data

Raw participant GPS records include latitude, longitude, accuracy, and the time at which the coordinate was recorded. Latitude and longitude indicate the record's location on a geographic coordinate system. Accuracy is determined based on the number of available satellites and their placement in the sky. It is defined as the radius of 68% confidence about the given location coordinates [48]. For example, if a coordinate has an accuracy of 10 meters then there is a 68% probability that the true location lies within 10 meters of the given location. These variables alone are insufficient to draw inferences about participant behaviors and require post-processing to extract meaningful data on periods of movement, travel mode, and significant locations visited [29]. Geofencing and cluster-detection approaches are two commonly used approaches to infer visited locations from participant GPS data [31,43,49]. Geofencing requires that the researcher know the location for each feature of interest (e.g. health facility) where visits are to be detected. A boundary is then defined around each feature, which may take the form of the building's footprint or may be a radius about a specified point of interest. Participant GPS data are then examined for sequential points located within the specified boundary, classifying those points as a visit when the number of points or the duration between the first and last point meet some predefined threshold. Several cluster-based approaches exist for visit detection with substantial variability in their methods [50-53] One example, ST-DBSCAN, identifies clusters according to three parameters: a minimum number of points, and a maximum spatial and non-spatial (e.g. time) distance that these points are located from one another [54]. Cluster-based approaches are more flexible than geofencing approaches in that they do not require any knowledge of where features are located; however, they tend to require larger datasets and consequently increased computing power.

Both approaches require user-specified parameters, optimal values for which vary with study aim, setting, and the mechanics of the specific approach being applied. Overly conservative parameters will fail to detect true visits while overly permissive parameters will falsely classify non-visits as visits. Studies detecting visits to general locations have generally applied duration values of 20 to 30 minutes [31,41,55,56], though at least one study applied a threshold of 5 minutes [57]. In contrast, a study identifying hospitalizations set a higher duration threshold of 4 hours [43]. There is less convergence around an optimal distance threshold with even those studies using similar duration parameters applying distance values ranging from 10 to 200 meters [41,56]. While several studies evaluate the performance of their approach after the fact, the specification of parameter values is typically based on *a priori* assumptions of optimal thresholds. A notable exception to this, Theirry et al. evaluated the performance of six parameter sets on 750 simulated trajectories to determine the optimal values [57]. Such exercises require a calibrating dataset where the true value is known for each record, which may not be feasible in all contexts. Given the sensitivity of visit detection approaches to parameter specification, the value of undertaking such a calibration exercise should not be overlooked.

Algorithms used to passively detect visited locations have struggled to balance detection of all true visits while minimizing the detection of false visits. There are limited studies applying a GPS-based approach to the detection of health facility visits. Nguyen and colleagues developed a GPS-based approach to detect hospitalizations across the United States, requiring that participants locate within a hospital's perimeter for at least four hours to be classified as a probable health facility visit [43]. Setting this high duration threshold reduced the probability of misclassifying shorter visits as hospitalizations and contributed to the overall correct detection rate of 65% [43]. While not explicitly targeting health facility visits, Paz-Soldan and colleagues applied a GPS-based approach to the detection of locations visits as part of a dengue transmission study in Iquitos, [31]. They provided participants with GPS data loggers, wearable GPS devices, and prospectively tracked their location for 14 days. The authors analyzed the collected GPS coordinates using a rule-based algorithm to determine visited locations. Participants then responded to a semi-structured interview (SSI) to identify all locations visited during the previous two weeks. The authors assessed overall concordance between the two methods and assessed concordance by the type of location detected, including 'health' locations. The authors also assessed concordance while varying several conditions of the detection algorithm. The GPS and SSI methods had perfect concordance when identifying the participant's home, 80-100% concordance when identifying schools and lodgings, 50-80% concordance when identifying residences, commercial locations, and religious locations. Concordance for the detection of health locations varied between 30-60% depending on the specifications of the detection algorithm. No information is provided on why performance was especially low for detecting visits to health facilities.

Continued development of GPS-based approaches to identify health facility visits seems likely. Earlier this year, Liss and colleagues published the results of formative research into acceptability of a smartphone-based tracking system to detect visits to the emergency department among high-risk adults and coordinate this information with patients' primary care providers, finding that patients were generally willing to utilize such a system if it improved the coordination of care services [58].

References

1 Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet (London, England). 2016;388:3027-35.

2 Jones G, Steketee RW, Black RE, Bhutta ZA, Morris SS. How many child deaths can we prevent this year? The Lancet. 2003;362:65-71.

3 Waldman RJ, Bartlett AV, Campbell CC, Steketee RW. Overcoming remaining barriers: the pathway to survival. Arlington, VA: BASICS Project, 1996.

4 Andersen RM. Families' Use of Health Services: A Behavioral Model of Predisposing, Enabling, and Need Components. Ann Arbor: Purdue University; 1968.

5 Andersen R, Newman JF. Societal and individual determinants of medical care utilization in the United States. The Milbank Memorial Fund quarterly Health and society. 1973;51:95-124.

6 Aday LA, Andersen R. A framework for the study of access to medical care. Health services research. 1974;9:208-20.

7 Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? Journal of health and social behavior. 1995;36:1-10.

8 Kroeger A. Anthropological and socio-medical health care research in developing countries. Soc Sci Med. 1983;17:147-61.

9 Bennett A, Eisele T, Keating J, Yukich J. Global trends in care seeking and access to diagnosis and treatment of childhood illnesses. Rockville, Maryland, USA: ICF International, 2015.

10 Geldsetzer P, Williams TC, Kirolos A, Mitchell S, Ratcliffe LA, Kohli-Lynch MK, et al. The recognition of and care seeking behaviour for childhood illness in developing countries: a systematic review. PLoS One. 2014;9:e93427.

11 Higgs G. A Literature Review of the Use of GIS-Based Measures of Access to Health Care Services. Health Services and Outcomes Research Methodology. 2004;5:119-39.

12 Guagliardo MF. Spatial accessibility of primary care: concepts, methods and challenges. Int J Health Geogr. 2004;3:3.

13 Nesbitt RC, Gabrysch S, Laub A, Soremekun S, Manu A, Kirkwood BR, et al. Methods to measure potential spatial access to delivery care in low- and middle-income countries: a case study in rural Ghana. Int J Health Geogr. 2014;13:25.

14 Costa LS, Nassi CD, Pinheiro RS, Almeida RM. Accessibility of selected hospitals and medical procedures by means of aerial and transit network-based measures. Health services management research. 2003;16:136-40.

15 Noor AM, Amin AA, Gething PW, Atkinson PM, Hay SI, Snow RW. Modelling distances travelled to government health services in Kenya. Trop Med Int Health. 2006;11:188-96.

16 Okwaraji YB, Cousens S, Berhane Y, Mulholland K, Edmond K. Effect of geographical access to health facilities on child mortality in rural Ethiopia: a community based cross sectional study. PLoS One. 2012;7:e33564.

17 Baume C, Helitzer D, Kachur SP. Patterns of care for childhood malaria in Zambia. Soc Sci Med. 2000;51:1491-503.

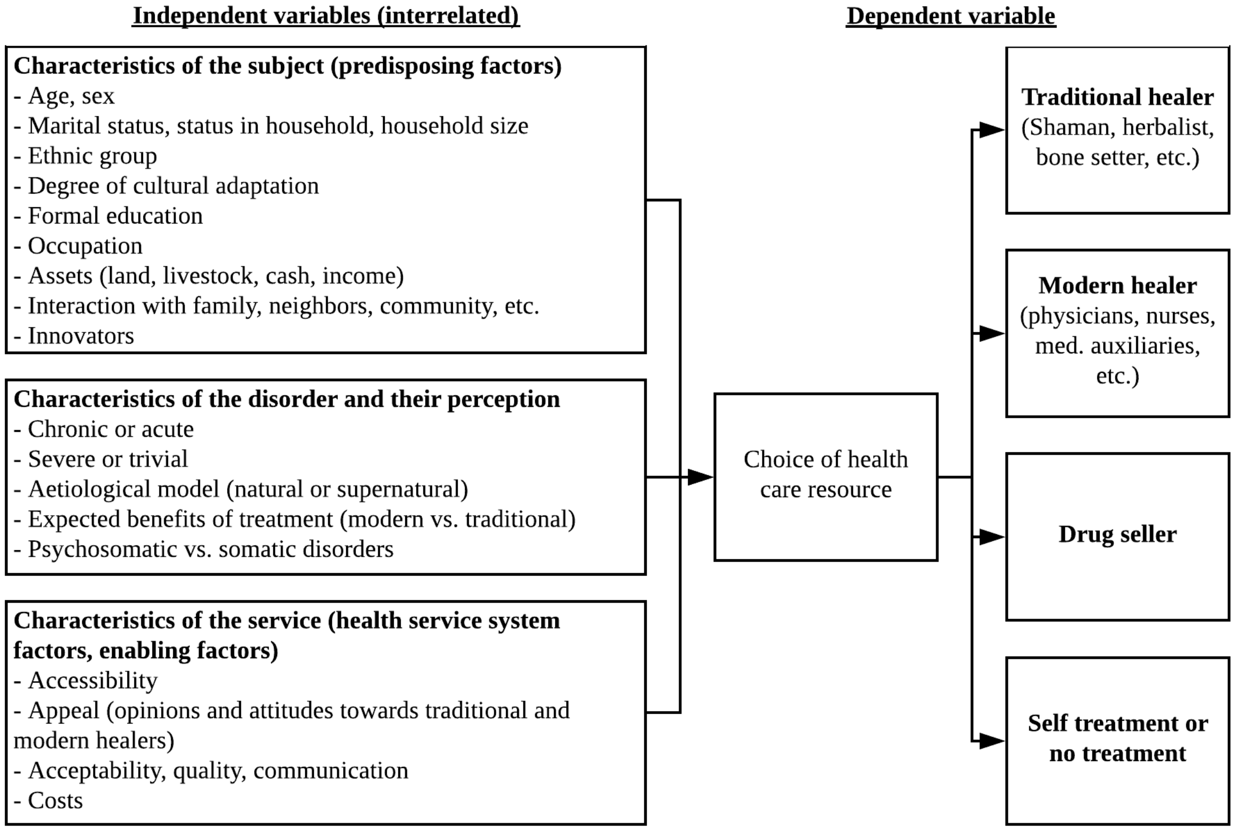
18 Bigogo G, Audi A, Aura B, Aol G, Breiman RF, Feikin DR. Health-seeking patterns among participants of population-based morbidity surveillance in rural western Kenya: implications for calculating disease rates. Int J Infect Dis. 2010;14:e967-73.
19 Wilson SE, Ouedraogo CT, Prince L, Ouedraogo A, Hess SY, Rouamba N, et al. Caregiver recognition of childhood diarrhea, care seeking behaviors and home treatment practices in rural Burkina Faso: a cross-sectional survey. PLoS One. 2012;7:e33273.
20 Glik DC, Ward WB, Gordon A, Haba F. Malaria treatment practices among mothers in Guinea. Journal of health and social behavior. 1989;30:421-35.
21 Gombojav N, Manaseki-Holland S, Pollock J, Henderson AJ. The effects of social variables on symptom recognition and medical care seeking behaviour for acute respiratory infections in infants in urban Mongolia. Arch Dis Child. 2009;94:849-54.
22 Das A, Ravindran TS. Factors affecting treatment-seeking for febrile illness in a malaria endemic block in Boudh district, Orissa, India: policy implications for malaria control. Malar J. 2010;9:377.
23 Chaturvedi HK, Mahanta J, Pandey A. Treatment-seeking for febrile illness in north-east India: an epidemiological study in the malaria endemic zone. Malar J. 2009;8:301.
24 Gao W, Dang S, Yan H, Wang D. Care-seeking pattern for diarrhea among children under 36 months old in rural western China. PLoS One. 2012;7:e43103.
25 Feikin DR, Nguyen LM, Adazu K, Ombok M, Audi A, Slutsker L, et al. The impact of distance of residence from a peripheral health facility on pediatric health utilisation in rural western Kenya. Trop Med Int Health. 2009;14:54-61.
26 Moisi JC, Kabuka J, Mitingi D, Levine OS, Scott JA. Spatial and socio-demographic predictors of time-to-immunization in a rural area in Kenya: Is equity attainable? Vaccine. 2010;28:5725-30.
27 Kroeger A. Response errors and other problems of health interview surveys in developing countries. World health statistics quarterly Rapport trimestriel de statistiques sanitaires mondiales. 1985;38:15-37.
28 Maas J, Sterkenburg RP, de Vries SI, Pierik F. Using GPS to Measure the Interaction Between Individuals and Their Neighbourhood. Neighbourhood Structure and Health Promotion2013. p. 153-75.
29 Kerr J, Duncan S, Schipperijn J. Using global positioning systems in health research: a practical approach to data collection and processing. Am J Prev Med. 2011;41:532-40.
30 Enge P, Misra P. Special issue on Global Positioning System. Proceedings of the Ieee. 1999;87:3-15.
31 Paz-Soldan VA, Reiner RC, Jr., Morrison AC, Stoddard ST, Kitron U, Scott TW, et al. Strengths and weaknesses of Global Positioning System (GPS) data-loggers and semi-structured interviews for capturing fine-scale human mobility: findings from Iquitos, Peru. PLoS Negl Trop Dis. 2014;8:e2888.
32 Michael Y, McGregor EM, Allen J, Fickas S. Observing Outdoor Activity Using Global Positioning System-Enabled Cell Phones. In: Helal S, Mitra S, Wong J, Chang CK, Mokhtari M, editors. Smart Homes and Health Telematics: 6th International Conference, ICOST 2008 Ames, IA, USA, June 28-July 2, 2008 Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 177-84.

33 Wu J, Jiang C, Liu Z, Houston D, Jaimes G, McConnell R. Performances of different global positioning system devices for time-location tracking in air pollution epidemiological studies. Environ Health Insights. 2010;4:93-108.
34 Zandbergen PA, Barbeau SJ. Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones. Journal of Navigation. 2011;64:381-99.
35 Benson AC, Bruce L, Gordon BA. Reliability and validity of a GPS-enabled iPhone "app" to measure physical activity. Journal of sports sciences. 2015;33:1421-8.
36 Garnett R, Stewart R. Comparison of GPS units and mobile Apple GPS capabilities in an urban landscape. Cartography and Geographic Information Science. 2015;42:1-8.
37 Jahkola O, Happonen A, Knutas A, Ikonen J, editors. What should application developers understand about mobile phone position data. ACM International Conference Proceeding Series; 2017.
38 Wu LL, Yang B, Jing P. Travel Mode Detection Based on GPS Raw Data Collected by Smartphones: A Systematic Review of the Existing Methodologies. Information. 2016;7:67.
39 Kazagli E, Chen J, Bierlaire M. Individual mobility analysis using smartphone data. Mobile Technologies for Activity-Travel Data Collection and Analysis2014. p. 187-208.
40 Wan N, Lin G. Life-space characterization from cellular telephone collected GPS data. Computers Environment and Urban Systems. 2013;39:63-70.
41 Brusilovskiy E, Klein LA, Salzer MS. Using global positioning systems to study health-related mobility and participation. Social Science & Medicine. 2016;161:134-42.
42 Glasgow ML, Rudra CB, Yoo EH, Demirbas M, Merriman J, Nayak P, et al. Using smartphones to collect time-activity data for long-term personal-level air pollution exposure assessment. J Expo Sci Environ Epidemiol. 2016;26:356-64.
43 Nguyen KT, Olgin JE, Pletcher MJ, Ng M, Kaye L, Moturu S, et al. Smartphone-Based Geofencing to Ascertain Hospitalizations. Circulation Cardiovascular quality and outcomes. 2017;10.
44 Palmer JR, Espenshade TJ, Bartumeus F, Chung CY, Ozgencil NE, Li K. New approaches to human mobility: using mobile phones for demographic research. Demography. 2013;50:1105-28.
45 Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res. 2013;15:e86.
46 Krenn PJ, Titze S, Oja P, Jones A, Ogilvie D. Use of global positioning systems to study physical activity and the environment: a systematic review. Am J Prev Med. 2011;41:508-15.
47 Bierlaire M, Chen JM, Newman J. A probabilistic map matching method for smartphone GPS data. Transportation Research Part C-Emerging Technologies. 2013;26:78-98.
48 Android Developers. Location | Android Developer. Available: https://developer.android.com/reference/kotlin/android/location/Location. Accessed: September 7, 2018.
49 Mazimpaka JD, Timpf S. Trajectory data mining: A review of methods and applications. Journal of Spatial Information Science. 2016;2016:61-99.
50 Xu R, Wunsch D, 2nd. Survey of clustering algorithms. IEEE Trans Neural Netw. 2005;16:645-78.

51 Palma AT, Bogorny V, Kuijpers B, Alvares LO. A clustering-based approach for discovering interesting places in trajectories. Proceedings of the 2008 ACM symposium on Applied computing - SAC '08; Fortaleza, Ceara, Brazil. 1363886: ACM; 2008. p. 863-8.
52 Yuan G, Sun PH, Zhao J, Li DX, Wang CW. A review of moving object trajectory clustering algorithms. Artificial Intelligence Review. 2017;47:123-44.
53 Zheng Y. Trajectory Data Mining: An Overview. Acm Transactions on Intelligent Systems and Technology. 2015;6.
54 Birant D, Kut A. ST-DBSCAN: An algorithm for clustering spatial–temporal data. Data & Knowledge Engineering. 2007;60:208-21.
55 Donaire-Gonzalez D, Valentin A, de Nazelle A, Ambros A, Carrasco-Turigas G, Seto E, et al. Benefits of Mobile Phone Technology for Personal Environmental Monitoring. JMIR Mhealth Uhealth. 2016;4:e126.
56 Sanchez M, Ambros A, Salmon M, Bhogadi S, Wilson RT, Kinra S, et al. Predictors of Daily Mobility of Adults in Peri-Urban South India. International journal of environmental research and public health. 2017;14:783.
57 Thierry B, Chaix B, Kestens Y. Detecting activity locations from raw GPS data: a novel kernel-based algorithm. Int J Health Geogr. 2013;12:14.
58 Liss DT, Serrano E, Wakeman J, Nowicki C, Buchanan DR, Cesan A, et al. "The Doctor Needs to Know": Acceptability of Smartphone Location Tracking for Care Coordination. JMIR Mhealth Uhealth. 2018;6:e112.

Tables and Figures

Figure 1.1 Andersen Behavioral Model as adapted by Kroeger



Source: [8]

Table 1.1 Selected studies evaluating travel impedance and utilization of child health services

Author, year	Study Design	Study Population	Exposure	Outcome	Effect estimate
Glik et al., 1989 [20]	Cross-sectional	Children 6-36 months; Guinea	Health center is perceived as near	Use of chloroquine during last fever episode	OR 1.9[1]
Baume et al., 2000 [17]	Cross-sectional	Children < 60 months; Zambia	Travel time to health center is less than one hour	Care seeking for fever and/or convulsions; returning for care when illness persists	Initial visit RR 1.36; Return visit RR 5.3[2]
Chaturvedi et al., 2009 [23]	Cross-sectional	General population; Assam, India	Distance to health center (≥5km vs. <5km)	Traditional healer as first source of care during episode of malaria	RR 1.4 (1.01,1.9)
Gombojav et al., 2009[21]	Prospective cohort	Children < 7 months; urban Mongolia	Distance to family's clinic (≥ 1km vs. < 1km)	Delayed care seeking for ARI (3+ days)	OR 3.52 (1.22, 10.18)
Feikin et al., 2009[25]	Cross-sectional	Children < 60 months; rural western Kenya	Distance from health facility	Incidence rates of pediatric clinic visits	IRR 0.66 (0.63,0.69) per 1km increase in distance
Bigogo et al., 2010[18]	Prospective cohort	General population; rural western Kenya	Distance from referral hospital	Care seeking for ARI, ALRI, or diarrheal disease from the local referral hospital	RR 0.821 (0.817, 0.826) per 0.5 km increase in distance
Das et al., 2010[22]	Cross-sectional	General population; Orissa, India	Distance traveled to seek care (≥5km vs. <5km)	Delayed (>48 hour) and/or ineffective treatment (no anti-malarial drug provided)	OR 2.04 (1.09, 3.83)
Moise et al., 2010[26]	Longitudinal	Children aged 1-2 months at enrollment; Kenya	Pedestrian and vehicular travel time to vaccination center	Time-to-immunization	Non-significant results
Gao et al., 2012[24]	Cross-sectional	Children <36 months; western China	Distance from the village to the township	Level of care sought for an episode of diarrhea (e.g. 0 = No care; 5 = County-level care)	"Population average effect" of - 0.09 (-0.18, - 0.01) per 10km increase in distance
Wilson et al., 2012[19]	Cross-sectional	Children <27 months; rural Burkina Faso	Distance from child's village to nearest public health center	Any care seeking outside the home for clinically defined diarrhea	OR 0.59 (0.42, 0.83) comparing children >10km from village to those located in the same village as the facility
Bennet et al., 2015[9]	Cross-sectional	Children <60 months, Zambia and Malawi	Distance of village center to nearest health facility (Farther than 5.7km [reference]; 3.6-5.7km [Group A]; 1.9-3.5km [Group B]; less than 1.9km [Group C])	Care seeking for fever Care seeking for diarrhea Care seeking for symptoms of ARI	Fever: Group A: 1.3 (1.1,1.5); B: 1.5 (1.3,1.8); C: 1.6 (1.4,1.9) Diarrhea: Group A: 1.1 (1.1,1.5); B: N/S; C: N/S Symptoms of ARI: Group A: 1.2 (1.0,1.5) B: 1.3 (1.0,1.6) C: 1.7 (1.3,2.2)

Note: OR = odds ratio; RR = relative risk; km = kilometer; ARI = acute respiratory infection; IRR = incidence rate ratio; ALRI = acute lower respiratory infection

[1] No confidence interval provided; the result is significant when adjusting for residence, but loses significance after accounting for the complete set of covariates

[2] The authors provided probabilities of each outcome among the exposed and unexposed, which we used to calculate the relative risk for each outcome

Chapter 2: Thesis Objectives

Objective 1: To evaluate the association between care-seeking for childhood cough, fever, or diarrhea and child, maternal, and household factors

Background:

An estimated 1.2 million children under five years of age die each year in India, with pneumonia and diarrhea among the leading causes. Increasing care-seeking is important to reduce mortality and morbidity from these causes. This paper explores the determinants and patterns of care-seeking for childhood illness in rural Pune district, India.

Methods:

Mothers having at least one child < 5 years from the study area of the Vadu Health and Demographic Surveillance System were enrolled in a prospective cohort study. Household sociodemographic information was collected through a baseline questionnaire administered at enrollment. Participants were visited up to six times between July 2015 and February 2016 to collect information on recent childhood acute illness and associated care-seeking behavior. Multivariate logistic regression explored the associations between care-seeking and child, participant, and household characteristics.

Findings:

We enrolled 743 mothers with 1,066 eligible children, completing 2,585 follow-up interviews (90% completion). Overall acute illness prevalence in children was 26% with care sought from a health facility during 71% of episodes. Multivariable logistic regression showed care-seeking was highly associated with the number of reported symptoms (adjusted odds ratio [AOR] = 2.4; 95% confidence interval [CI] 1.5-3.9; $p<0.01$) and household insurance coverage (AOR = 2.2; 95% CI 1.1-4.3; $p<0.05$). We observed an interaction between the associations of illness severity and maternal employment on care-seeking. Moderate-to-very severe illness was associated with increased care-seeking among both employed (AOR 5.0; 95% CI 2.2-11.1; $p<0.001$) and currently unemployed mothers (AOR = 7.0; 95% CI 3.9-12.6; $p<0.001$). Maternal employment was associated with reduced care-seeking for non-severe illness (AOR 0.3, 95% CI 0.1-0.7; $p<0.001$), but not associated with care-seeking for moderate-to-very severe illness. Child sex was not associated with care-seeking.

Conclusions:

This study demonstrates the importance of illness characteristics in determining facility-based care-seeking while also suggesting that maternal employment resulted in decreased care-seeking among non-severe illness episodes. The nature of the association between maternal employment and care-seeking is unclear and should be explored through additional studies. Similarly, the absence of male bias in care-seeking should be examined to assess for potential bias at other stages in the management of childhood illness.

Objective 2: To evaluate the feasibility of a smartphone-based approach for tracking participant movement and to explore factors associated with the approach's success

Background:

Common approaches to measure health behavior depend on the validity of participant responses and may be biased due to poor recall or non-completion. Technology-based approaches, such as those using Global Positioning System (GPS) data, provide an alternative approach while reducing these biases. This paper describes the development and implementation of TrackCare, a location-aware smartphone application used to detect maternal care-seeking for childhood illness.

Methods:

Mothers having at least one child < 5 years were enrolled and provided a GPS-enabled smartphone preinstalled with TrackCare, which recorded device location each minute and transferred data hourly to a central server. Household sociodemographic information was collected at baseline and field workers visited mothers monthly over six months to assess compliance with phone-related procedures. Location data completeness and quality were evaluated prospectively. Separate analyses evaluated the associations between various explanatory variables, data completeness, and participant compliance with study procedures, respectively.

Findings:

We enrolled 200 mothers and completed 1,092 follow-up interviews. Participant smartphones submitted location data for a mean of 152 days, representing 84% data completeness across the study duration. Compliance with study phone-related study procedures was 79% overall and significantly associated with increased data completeness ($\boldsymbol{\beta}_{\boldsymbol{Visits\,1-3}}$ = 8%; 95% confidence interval [CI] 3-13%; $\boldsymbol{\beta}_{\boldsymbol{Visits\,4-6}}$ = 19%; 95% CI 10-28%). Urban residence was associated with decreased data completeness ($\boldsymbol{\beta}$ = -4%; 95% CI -9% – -0.3%). Compliance was highest among participants of higher socioeconomic status (adjusted odds ratio [AOR] = 1.67; 95% CI 1.13 – 2.46) and increased during the second half of the study (AOR = 1.66; 95% CI 1.23 – 2.24). No association was observed between compliance and maternal age, education, baseline perceptions, or previous smartphone ownership.

Conclusions:

GPS-enabled smartphones provide a promising approach for the continuous, real-time measurement of movement data with increased data completeness relative to traditional GPS devices**.** These data provide useful insights into a range of participant health behaviors, such as care-seeking for childhood illness, with the potential to enhance data collection methods as approaches to analyze these location data are refined.

Objective 3: To evaluate the correct detection rate of a GPS-based method to detect health facility visits and explore the factors associated with the improved performance of the method

Background:

Common approaches to measure health behaviors rely on the validity of participant responses and are subject to bias. Technology-based alternatives, particularly those using GPS, address these biases while opening new channels for research. This study describes the development and implementation of a GPS-based approach to detect health facility visits in rural Pune district, India, nested within a study evaluating maternally reported care-seeking behavior for childhood illness.

Methods:

Participants were mothers of children under-five within the study area of the Vadu Health and Demographic Surveillance System. Participants received GPS-enabled smartphones preinstalled with a location-aware application, TrackCare, that continually recorded participant GPS data and transferred these data to a central server. Data were analyzed to identify health facility visits according to a parameter-based approach, optimal values of which were calibrated through locally simulated health facility visits. Participants reviewed all detected visits through monthly prompted recall surveys, confirming those visits which were correctly identified. Detected visits were analyzed using logistic regression methods to explore factors associated with the detection of false positive visits.

Findings:

Field workers enrolled 200 participants and completed 1,098 follow-up visits over the six-month study period. Prompted recall surveys were completed for 694 follow-up visits with at least one GPS-based health facility visit detected. While the approach performed well during calibration (AUC 0.95), overall performance was low when applied to participant data with 440 of 22,251 detected visits confirmed (2%). False positives increased as participants spent more time in high health facility density areas (odds ratio [OR] 2.29, 95% confidence interval [CI] 1.62-3.25). Visits detected at facilities other than hospitals and clinics were also more likely to be false positives (OR 2.78, 95% CI 1.65-4.67). False positives were more likely among visits detected nearby participant homes with the likelihood decreasing as distance increased (OR 0.89, 95% CI 0.82-0.97). Visit duration was not associated with confirmation status.

Conclusions:

Field-worker simulated health facility visits resulted in an optimal parameter combination that performed well within the simulation dataset but substantially overestimated health visits when applied to participant GPS data. This study provides useful insight into the challenges in detecting health facility visits where providers are numerous and their location is highly clustered within urban centers.

Chapter 3: Methods

This thesis research was nested within a prospective cohort study exploring the validity of maternal recall of care-seeking for childhood illness, conducted by the Vadu Health and Demographic Surveillance System (HDSS) in partnership with the University of Edinburgh (UoE) and the Institute for International Programs at Johns Hopkins University (IIP-JHU). The study was one of several conducted within the Improving Coverage Measurement in Maternal, Newborn, and Child Health (ICM) project, funded through a grant from the Bill and Melinda Gates Foundation (BMGF).

3.1 Methods for Improving Coverage Measurement – India Project

From November 2014 – February 2016, the Vadu HDSS conducted a validation study in collaboration with UoE and IIP-JHU under the direction of principal investigators Sanjay Juvekar (Vadu HDSS) and Harry Campbell (UoE). The study was funded by a prime award by BMGF made to IIP-JHU of which UoE and Vadu HDSS were sub-grantees and was designed to explore the validity of maternally-reported care-seeking for childhood illness among mothers of young children in rural Pune district, Maharashtra state, India. The design included a GPS-based approach to define care-seeking events, serving as the comparator against which maternally-reported care-seeking were evaluated. A separate study within the ICM project evaluated the validity of maternally-reported care-seeking in Zambia through several alternative approaches [59]. The ICM India study was unique among the studies within the larger ICM project in its application of a GPS-based approach for comparison with maternally-reported care-seeking.

3.1.1 Study Aims and Objectives

The aim of the study was to compare reports of care seeking obtained in response to care-seeking questions in a modified Demographic and Health Survey (DHS) questionnaire with "actual" care-seeking events determined through a GPS-based approach. This approach used location services built into a smartphone application to determine when a care-seeking visit has occurred. These visits constituted the operational gold standard for defining "actual" care-seeking events. The primary objectives of the ICM India study were:

1. To estimate what percent of all reported care-seeking events for children under five years of age actually took place
2. To estimate what percent of all "actual" care-seeking contacts for children under five years of age are reported by caregivers

3.1.2 Study Population

The ICM India took place in the 22 contiguous villages of the Vadu HDSS study area (Figure 3.1.2.1), located about 30 kilometers northeast of the Pune metro area. These villages represented a subset of two administrative blocks in rural Pune district, Shirur and Haveli, and covered a total area of approximately 240 square kilometers. Population density was highest in four urban villages situated along the main road that connects Pune city with the city of Shirur, 70 kilometers further northeast. These urban villages also included many large buildings that may have interfered with GPS signal quality. The 2011 Census of India provided the following information on the study area: adult literacy was 85% with literacy higher among men than among women (91% vs. 79%);

the child sex ratio favored boys, with approximately 1.17 boys aged 0-6 years per each girl; 46% of the working population was employed in agriculture; and approximately 12% of the population belonged to either a scheduled caste or scheduled tribe [60]. Several public health facilities were located in the study area, including a rural hospital, primary health center, and several sub-centers, but the majority of facilities were in the private sector, which included a number of private clinics, smaller hospitals, and one NGO hospital.

Data on the epidemiology of childhood illnesses and care seeking behaviors are available at the state level through the NFHS, the most recent iteration of which reported a two-week prevalence of symptoms of ARI, fever, and diarrhea in Maharashtra of 2%, 13%, and 9% with care being sought for each illness in 89%, 85%, and 78% percent of episodes, respectively [61]. The Public Health Foundation of India (PHFI) and UNICEF recently conducted reviews addressing newborn and child health, including reviews specifically targeting ARI and diarrheal disease [62,63]. These reviews provide insight into the care-seeking practices for these illnesses, revealing considerable heterogeneity among care-seeking practices. The review of care seeking practices for ARI presents the results of the third round of the NFHS, finding that at the national level urban residence, higher education, and higher socioeconomic status were associated with increased care-seeking [62]. Distance to the health facility was listed among the barriers to seeking care at a health facility as well as the related factors of travel time and travel cost [62]. The review of care seeking for diarrheal disease identifies similar determinants at the national level as for ARI while also noting an association between male sex and increased likelihood of care-seeking [63]. Additionally, these reviews highlight the differences in care seeking practices by region, comparing the southern state of Kerala where care-seeking was high and most care was sought from western providers with studies from the northern city of Lucknow where traditional healers were consulted for a quarter of neonatal ARI and a third of neonatal diarrhea and the northern state of Rajasthan where two thirds of sick newborns were not brought for care [64-66].

Nationally, care is most frequently sought from private sector practitioners and this increases proportionally to household socioeconomic status [67]. While the private sector remains the most common source of care within the state of Maharashtra, a mixed-methods study from a rural region found high use of traditional healers for selected newborn danger signs (poor sucking, difficult breathing, and boils over body) and symptoms of childhood illness (measles) [68]. Separately, there is evidence of male bias in the management of childhood illness from experience within the study area. A study conducted in rural Pune district identified that boys were more likely to be brought to private providers, more likely to comply with referral, tended to have more money spent on treatment, and were brought further from home for care [69]. The most recent state-level data do not suggest male bias in the proportion of children brought to a health facility for care [61], though this does not rule out the possibility of bias at other stages in the management of childhood illness.

3.1.3 Study Participants

Mothers

The study included 749 mothers ages 15-49 with at least one child age 0-59 months. These age criteria were established to reflect the criteria used by DHS, implemented within India as the NFHS. We also required that the mother self-identify as the primary care giver for the eligible children. Mothers were randomly sampled from a population register maintained by the Vadu

HDSS. Field staff approached mothers at their homes, provided information on the study in the local language (Marathi), and asked mothers for informed consent.

The phone group included 200 mothers while the comparison groups included a total of 549 mothers. Of these 549 mothers, 100 were in the longitudinal comparison group and 449 were in the cross-sectional comparison group, further subdivided into six groups of about 75 mothers each. All eligible children (i.e. <60 months) for each enrolled mother were included in the study.

The sample size for the phone group was determined based on an estimated two-week period prevalence of care-seeking of 20%, an average of two eligible children per enrolled mother, a base prevalence of accurate care-seeking of 80% and a precision level of 8%. The sample size for the comparison groups was determined based on what was considered feasible given available resources. Figure 3.1.3.1 illustrates the follow-up schedule for each study group.

Health Facilities

The study included all health facilities identified in the study area. The term "health facility" was broadly applied to include any public, private, formal, and informal location where a mother may seek advice or treatment for an episode of childhood illness. Health facilities were identified by field staff with experience in the study area. It was possible that a facility was missed during this census or that a new facility opened during the course of the study. We included two mechanisms to identify these additional facilities through the course of the study. In the baseline questionnaire we asked mothers the names of facilities typically visited during an episode of childhood illness and in the follow up questionnaire we ask the name of the facility actually visited during the recent episode of childhood illness (whenever applicable). We checked these facilities against the list generated during the health facility census to identify any additional facilities. A previously conducted census of the formal health sector suggested that up to 150 formal facilities might operate in the study area. No such estimate was available for facilities in the informal health sector. The health facility questionnaire was administered to those facilities providing informed consent. The location of each health facility was considered to be part of the public domain. Consent was therefore not sought prior to collecting GPS coordinates.

Recorded GPS coordinates were imported to Google Earth and superimposed on available satellite imagery. The coordinates were then reviewed in collaboration with local staff to determine their accuracy. Any coordinates appearing to be more than 5 meters from the health facility entrance according to the satellite image were collected again. Coordinates that repeatedly appeared to be more than five meters from the health facility were manually adjusted.

3.1.4 Study Design and Procedures

This was a prospective validation study of mothers' responses to standard DHS questions about care-seeking behaviors. To validate mothers' responses required the use of a gold standard method to assess actual care-seeking behavior. A true gold standard would require continuous, direct observation of mothers, which was not be feasible for several reasons. (e.g. privacy, cost) The study used an mHealth approach as an operational gold standard to define actual care-seeking behavior. This mHealth approach used GPS coordinates recorded from smart phones provided to 200 participants ("phone group") to assess actual care seeking behavior. A complete description of the development and evaluation of this mHealth method is included in Chapter 5. Participants were followed up monthly for six months and administered a questionnaire about care seeking for childhood illness in the previous two weeks. The reference period was selected to maximize comparability with the DHS and MICS questionnaire. These responses were then compared to care seeking detected through the mHealth approaches to determine the accuracy of reported care seeking behavior.

We identified two aspects of the study design that may potentially bias participant recall and/or reporting of care-seeking behavior. The first was the presence of the phone (Hawthorne effect) and the second was the administration of the follow-up questionnaire over six rounds (repeated testing effect). To assess for these effects, we included two comparison groups within our study. The first is a longitudinal comparison group that was followed up according to the same schedule as the phone group and the second is a set of six cross-sectional comparison groups that were each be followed up only once during the course of the study. Participants in neither comparison group were provided with a smartphone.

The study includes five data collection forms, all translated into the local Marathi. Table 3.1.4.1 presents a brief description of these forms.

3.1.5 Study Phone Selection and Testing

Participant location data were collected through a location-aware smartphone app installed on GPS-enabled smartphones. Details of the application and smartphones are provided below with further details provided in Chapter 5.

TrackCare Application

The TrackCare application was developed for the Android operating system, version 4.4, and used the built-in Android location services API to continuously monitor the phone's location. The app communicated with the operating system to receive location updates, the timing of which depended on quality of the phone's connection with location providers (e.g. GPS satellites). Each time an updated location was received, it was compared against the phone's current best estimate of its location. This comparison considered whether the updated location was newer and/or more accurate than the current location. These criteria were adopted based on the suggestions included in the Android Developer Location Strategies documentation [70]. If the location update was better, then it replaced the previous location as the current best estimate for the phone's location. If no location updates were received (e.g. due to poor connectivity), then the current best estimate was carried forward.

Each minute the phone's current best location estimate was saved in an internal database. Each record included: latitude, longitude, accuracy, data source (i.e. GPS or network provider), current location mode on device, and an indicator variable reporting whether the record represents a new or a previously obtained (i.e. cached) GPS coordinate (Table 3.1.5.1). Accuracy is defined as the radius of 68% confidence about the given location coordinates [48].

Recorded data were encrypted, stored locally on the device, and transmitted hourly to a central study server using the phone's network connection. Whenever hourly transfer was not possible (e.g. due to poor network connection), the app transmitted all cumulated points during the following scheduled transfer. Local data were deleted from the device after successfully being transferred to the central server.

The app required no interaction from participants and was programmed to run continually as a background service. The app launched automatically after installation. The app would re-launch whenever the phone was restarted or in the event that the app was forced quit by the participant. We also included a feature to designate the app as a device administrator, which when enabled prevented the participant from uninstalling the app. Device administrator status was enabled by default and a password was required to disable it. Uninstallation of the app by the participant remained possible by restoring factory settings, though we expected this would be a rare occurrence. Together, we expected these additional features to ensure that the app remained installed on the phone and operated continually as long as the phone was turned on.

A minimal user interface was designed to assist field workers with checking the status of the application and debugging any issues. The interface was organized around four tabs. These tabs presented the most recent location data obtained by the phone and displayed the number of records currently saved in the phone's database. Assuming that all data were successfully transferred each hour, the number of records saved on the phone should be 60 or fewer. High values for this field prompted field workers to troubleshoot for possible network connectivity issues.

Data were transferred from the phones to a study server located at the University of Edinburgh and a cloud-based mirror server maintained by the Vadu HDSS. Data stored on the cloud-based server were incrementally downloaded to a local server located at Vadu HDSS.

An automated Stata script processed the incremental data each day and checked the quality of data submitted by each participant. Data processing included removing duplicate records, assigning value labels to categorical fields, and appending the incremental data to the main dataset. Data quality checks examined the previous 14 days of participant and considered several indicators of quality. Additional information on these data quality checks is included below.

Phone Selection

We pilot tested several phone models and selected from among them the model best suited to the study. Piloted phones were selected based on price, local availability, and a variety of technical specifications (e.g. Android operating system, GPS chipset, internal storage capacity, battery capacity). The TrackCare app required a constant GPS connection and placed a substantial burden on the battery beyond that expected through normal use. In addition to reviewing the stated battery capacity of each phone, we also conducted field testing of the phones performance while running the TrackCare app. Additional testing assessed the quality of the individual GPS data collected by the phone relative to known locations, with a particular focus on the quality of data obtained during simulated health facility visits. The final model was selected based on the quality of recorded GPS data, battery life, local availability, and price.

Initial pilot testing supported the selection of the Sony Xperia E1 Dual Sim model, though due to unforeseen challenges in obtaining sufficient quantity of this model we selected the Sony Xperia E4 Dual Sim model. Prior to study initiation we compared the E4 and E1 models and the E4 model was observed to function at least as well as the E1 model. The 200 mothers enrolled during the main study were provided with the E4 model while the 10 enrolled from the pilot continued to use the E1 model that had been previously distributed to them.

Preparing Smartphones for Deployment

We configured all study phones at the field office prior to distribution. During configuration, we removed unnecessary pre-installed software, installed the TrackCare app, and ensured that optimal phone settings were activated. These included mobile data being enabled, Wi-Fi being turned on, and location services enabled with the location mode set to “High Accuracy.” We inserted a study SIM card into each study phone, which included a mobile data plan to enable the transfer of participant data to the study server via the available cellular network. We purchased all SIMs through the Idea Cellular service provider, which is reported by field staff to provide the most reliable network coverage within the study area. Vadu HDSS has an existing group plan with this provider that facilitated the purchase of the 210 SIMs required for the study. The group plan also provides free in-group calling, which facilitated communication between field staff and study participants.

We initially considered installing a free, third-party application that would lock specific phone settings as per our specifications, though pilot testing revealed that such apps could interfere with the normal functioning of the phone and could result in device instability.

Participant Compliance

We provided participants with a phone information leaflet at baseline, which was translated into Marathi and instructed mothers on the proper phone usage. Key usage instructions included charging the phone daily, keeping it on during the day, carrying it during any care-seeking event, and refraining from altering the phone settings. Throughout the course of the study, we monitored compliance through systematic and random assessments. The systematic components will further be divided into two separate activities. These are described briefly below.

- **Systematic Assessment of Compliance**
    - **Phone Check-up:** We administered a brief questionnaire to assess phone usage habits since the last study contact and identify any changes to the phone's settings that could have interfered with the study. This questionnaire was administered to all mothers once three days after enrollment, once 11-14 days after enrollment, and once at each monthly follow-up visit. These data provided the basis for assessing participant compliance with phone-related study procedures at each follow-up visit.
    - **Daily Data Checks:** Each day we analyzed the previous two weeks of submitted data for each participant. The assessment focused on the total quantity of points submitted, the percent of the total points located outside of the study area, and the percent of the total points for which the device location mode had been changed from its optimal setting. We established thresholds for the acceptable use and followed up with mothers failing to meet those thresholds through phone calls and home visits.
- **Random Assessment of Compliance:** Each day we called a random sample of 5% of enrolled mothers and administered a short questionnaire about phone usage habits during the previous day. Another aim of this activity was to assess whether the phone remained with the study participant. The sample for each day was selected without replacement but

the sampling over time was done with replacement. In other words, a mother would not be included twice on the same day but may have been included on two or more consecutive days.

Data Processing and Cleaning

The above steps resulted in the collection of a large quantity of raw GPS data, which underwent substantial processing prior to conducting analysis.

Incorrect Times

During data collection we observed that several phones were submitting coordinates with timestamps that were clearly incorrect. We first identified this problem by comparing the time when the record was saved on the phone (determined by phone's internal clock) and the time when the record was transferred to the study server (determined by server). Given the order that these steps occur, the time the record was saved on the phone should always have been before the time that the record was transferred to the server. We noticed that the opposite had occurred, which indicated that one of the time values was incorrect. Through further investigation we confirmed that the error was with the phone's internal clock, which artificially shifted forward by some increment of time. The internal clock would eventually correct, though all points saved during this time would be offset by the same amount of time. Observed increments were typically in multiples of whole days or whole hours, but non-standard increments were also observed. Detecting blocks of points for which the clock had artificially shifted forward was relatively simple, but we later discovered blocks for which the reverse had occurred (i.e. artificial shift back in time). An additional category of offset was detected when the phones' internal clock appeared to reset to January 1, 2010.

Points with offset times jeopardized our ability to draw a valid inference about participant movement. Due to variability in how these offsets presented themselves in the data, determining the magnitude of the problem required substantial effort. We applied a highly sensitive definition to identify all possible locations where any offset may have occurred and reviewed each case manually to assess for the presence of an offset. In total, we identified about 800,000 points (1.5%) for which the time was incorrect. Minimal incremental effort was required to correct these times once they were identified, so we opted to do so rather than drop them from the dataset.

We are not aware of what may have caused these offsets. While it is possible that the incorrect times could result from participant manipulation, our experience with the pattern of these offsets suggests this is not the case. We expect that this was related to some unexpected behavior at the level of the device or the operating system. Improved understanding of the problem would result in less intensive approaches to promptly detect and remedy such offsets.

Unreliable Location Data

The location data obtained through the TrackCare app included three main fields: latitude, longitude, and accuracy. The latitude and longitude values located the phone at a specific point on the map, while the accuracy value provided the radius of 68% confidence about that point with a smaller value representing a more precise location estimate [48]. Highly accurate GPS devices can routinely obtain coordinates with an accuracy below 10 meters in good conditions, while accuracy can degrade to 50 meters in poorer conditions [29]. Previous studies recommend dropping records where the accuracy value exceeds a predetermined threshold. Zhou and colleagues exclude records with accuracy greater than 35 meters [71], while Montini and colleagues adopted a marginally higher threshold of 50 meters [72]. We adopted the higher threshold of 50 meters in order to maximize the number of retained points. Points with an accuracy value above this threshold were excluded from analysis.

Missing Data

Records dropped as part of the above process resulted in gaps in the dataset. We also expected gaps during periods when the normal functioning of the app had been interrupted (e.g. due to phone being off). We adopted the approach put forward by Kestens and colleagues and interpolated between valid data points if the time between the two points was less than one hour or the distance was less than 100 meters [73]. If consecutive points were greater than one hour and 100 meters apart then we did not attempt interpolation and the intermediate points remained missing.

3.2 Methods for This Thesis Research

This thesis research used data collected as part as part of ICM India parent study. Informed consent for the parent study was obtained from study participants, mothers of children under five. The ethics committees of KEM Hospital Research Centre and the University of Edinburgh approved the study protocol for the parent study, which included all components addressed within this thesis research. Given these approvals, this thesis research did not require approval from the institutional review board of the Johns Hopkins Bloomberg School of Public Health. The original consent documents included permissions to conduct all elements of this thesis research, including the analysis of care-seeking determinants, the tracking of participant movement data, and the analysis of participant movement data to detect possible health facility visits. All participants included within the ICM India study were eligible for inclusion within this thesis research. Various subsets of the total ICM India study population were included depending on the thesis objective.

3.2.1 Objective 1 Methods

The first research objective was to evaluate the association between individual and household variables, including distance from a health facility, and care-seeking for childhood illnesses. Using a prospective cohort design, 749 mothers of children under 60 months of age were recruited from the 22 villages of the Vadu HDSS study area. Field workers enrolled the mothers and their eligible children during home visits and randomly assigned mothers to one of three study groups. At enrollment field workers collected information on individual and household factors associated with care seeking while also recording the GPS coordinates of the participant's home. Mothers were then followed up over six months according to group-specific schedules. At each follow-up visit, field workers collected information on childhood illness, subsequent care-seeking, and the location where care was sought (if applicable). Prior to study initiation, field workers conducted a census of all health facilities in the study area during which they recorded the GPS coordinates of each health facility and classified all facilities by type. Using participant household locations, facility locations, and the Vadu road network, we computed the network-based distance from each participant to each health facility within the study area, and identified the facility that is nearest to each participant. Supplementary information collected during follow-up identified the sequence of providers consulted for each illness episode.

3.2.2. Objective 2 Methods

The second objective was to evaluate the feasibility of a smartphone-based approach for tracking participant movement and to explore factors associated with the approach's success. Using a prospective cohort design, 200 mothers of children under 60 months of age were recruited from the 22 villages of the Vadu HDSS study area. Field workers enrolled participants during home visits and provided each participant with a GPS-enabled smartphone, preinstalled with a location-aware application ("TrackCare"). The Trackcare app recorded participant location each minute and transmitted these data to a central server for further analysis. Automated Stata scripts assessed data completeness throughout the study. Field workers provided support to participants as needed to address any data completeness issues. Raw GPS data were processed to remove unreliable data points and interpolate participant location during periods where these data were missing. At enrollment, field workers collected detailed information on socio-demographic factors and perceived acceptability of using mobile phones to monitor care seeking behavior. Field workers then assessed participant compliance with study procedures at each of six monthly follow-up visits.

3.2.3 Objective 3 Methods

The third objective was to evaluate the performance of a GPS-based method to detect health facility visits and explore the factors associated with the method's optimal performance. Using a prospective cohort design, 200 mothers of children under 60 months of age were recruited from the 22 villages of the Vadu HDSS study area. Field workers enrolled mothers during home visits. At enrollment field workers collected detailed information on socio-demographic factors and perceived acceptability of using mobile phones to monitor care seeking behavior. Field workers recorded the GPS coordinates of each mother's home. Field workers also conducted a census of health facilities in the study area, which included collected the facility coordinates and classifying each facility by type. Mothers were provided with GPS-enabled smartphone preinstalled with the TrackCare application, which prospectively recorded their movement data and transmitted these data to a central server for further analysis. A rule-based geofencing algorithm was developed to assess whether a set of points constitutes a probable visit to a health facility. Field workers visited mothers each month with a list of GPS-detected visits and reviewed these visits with mothers through prompted recall surveys to identify which visits were correctly detected and which were false positives.

3.4.4 Analytic Methods

Analytic methods are described for each objective in chapters 4-6.

References

29 Kerr J, Duncan S, Schipperijn J. Using global positioning systems in health research: a practical approach to data collection and processing. Am J Prev Med. 2011;41:532-40.

48 Android Developers. Location | Android Developer. Available: https://developer.android.com/reference/kotlin/android/location/Location. Accessed: September 7, 2018.

59 Carter ED, Ndhlovu M, Munos M, Nkhama E, Katz J, Eisele TP. Validity of maternal report of care-seeking for childhood illness. J Glob Health. 2018;8:010602.

60 Census of India. Administrative Atlas of India. New Delhi: Office of the Registrar General & Census Commissioner, 2011.

61 International Institute for Population Sciences (IIPS) and ICF. National Family Health Survey (NFHS-4), 2015-16: India. Mumbai: IIPS, 2017.

62 Mathew JL, Patwari AK, Gupta P, Shah D, Gera T, Gogia S, et al. Acute respiratory infection and pneumonia in India: a systematic review of literature for advocacy and action: UNICEF-PHFI series on newborn and child health, India. Indian pediatrics. 2011;48:191-218.

63 Shah D, Choudhury P, Gupta P, Mathew JL, Gera T, Gogia S, et al. Promoting appropriate management of diarrhea: a systematic review of literature for advocacy and action: UNICEF-PHFI series on newborn and child health, India. Indian pediatrics. 2012;49:627-49.

64 Pillai RK, Williams SV, Glick HA, Polsky D, Berlin JA, Lowe RA. Factors affecting decisions to seek treatment for sick children in Kerala, India. Soc Sci Med. 2003;57:783-90.

65 Awasthi S, Srivastava NM, Pant S. Symptom-specific care-seeking behavior for sick neonates among urban poor in Lucknow, Northern India. J Perinatol. 2008;28 Suppl 2:S69-75.

66 Mohan P, Iyengar SD, Agarwal K, Martines JC, Sen K. Care-seeking practices in rural Rajasthan: barriers and facilitating factors. J Perinatol. 2008;28 Suppl 2:S31-7.

67 Sreeramareddy CT, Sathyanarayana TN, Kumar HN. Utilization of health care services for childhood morbidity and associated factors in India: a national cross-sectional household survey. PLoS One. 2012;7:e51904.

68 Dongre AR, Deshmukh PR, Garg BS. Childhood morbidity, household practices and health care seeking for sick children in a tribal district of Maharashtra, India. Indian J Med Sci. 2010;64:7-16.

69 Ganatra B, Hirve S. Male bias in health care utilization for under-fives in a rural community in western India. Bull World Health Organ. 1994;72:101-4.

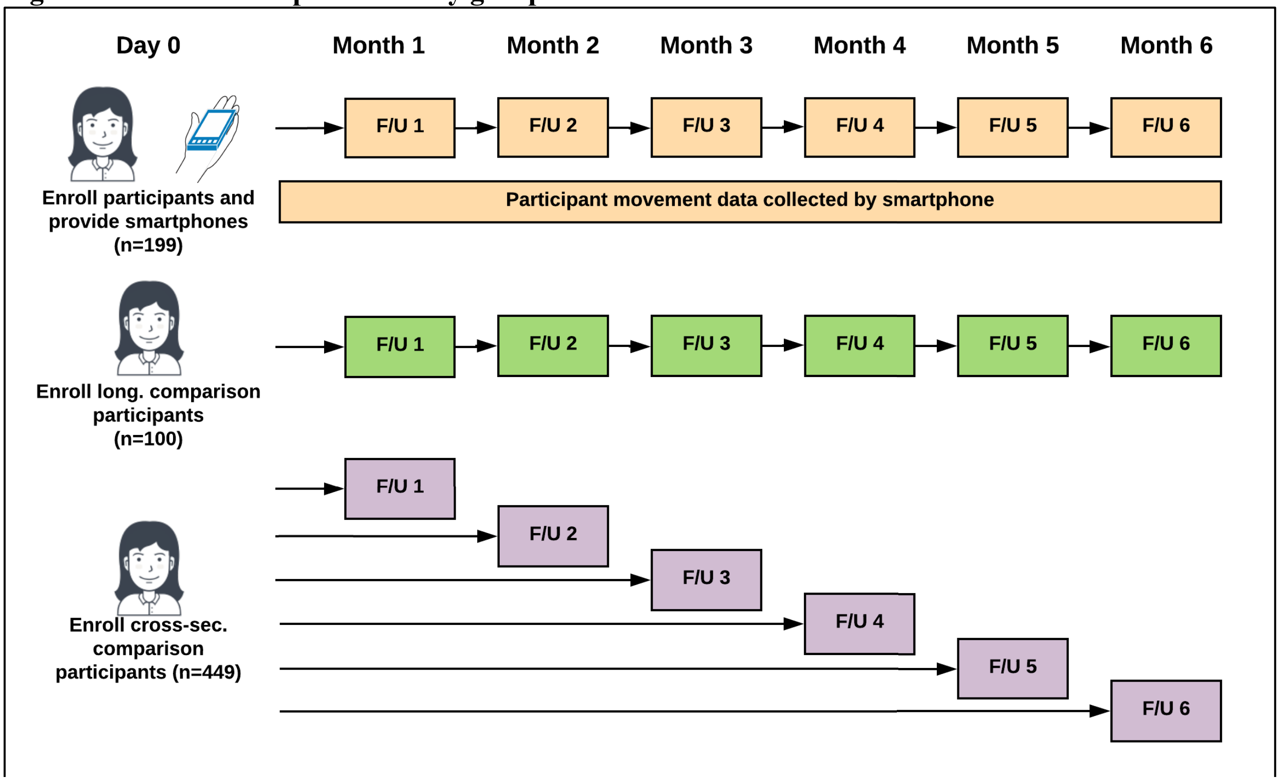
70 Google. Location Strategies | Android Developer. n.d. Available: https://developer.android.com/guide/topics/location/strategies.html. Accessed: June 11, 2017.

71 Zhou XL, Yu W, Sullivan WC. Making pervasive sensing possible: Effective travel mode sensing based on smartphones. Computers Environment and Urban Systems. 2016;58:52-9.

72 Montini L, Prost S, Schrammel J, Rieser-Schassler N, Axhausen KW. Comparison of travel diaries generated from smartphone data and dedicated GPS devices. Transport Survey Methods: Embracing Behavioural and Technological Changes. 2015;11:227-41.
73 Kestens Y, Thierry B, Chaix B. Re-creating daily mobility histories for health research from raw GPS tracks: Validation of a kernel-based algorithm using real-life data. Health Place. 2016;40:29-33.

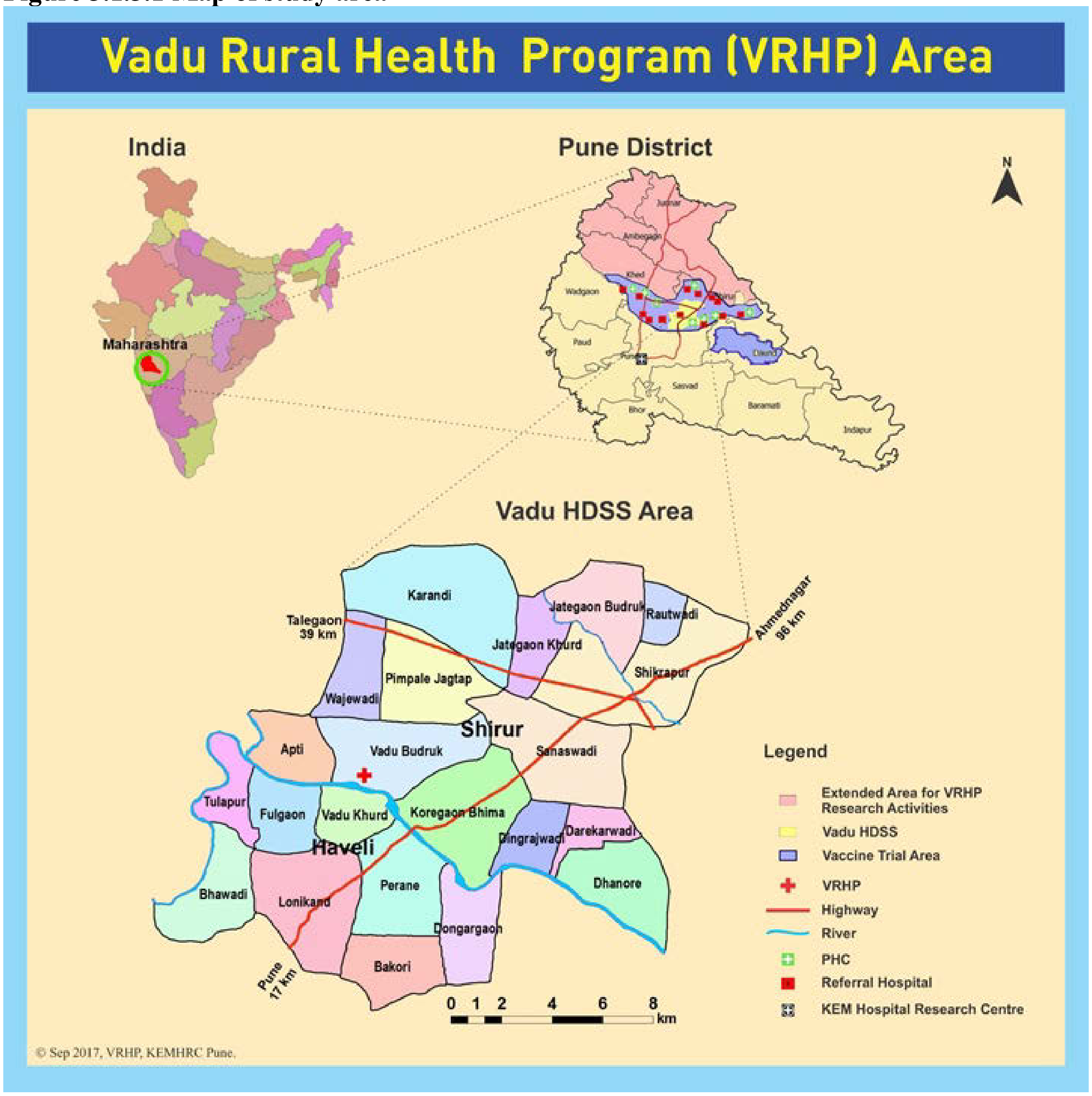
Tables and Figures

Figure 3.1.2.1 Follow-up schedule by group



Note: Long. comparison group = longitudinal comparison group; Cross-sec. comparison group = cross-sectional comparison group

Figure 3.1.3.1 Map of study area



Source: Vadu Rural Health Program, KEMHRC, Pune

Table 3.1.4.1 Data collection forms for Improving Coverage Measurement Study, India

#	Name	Group	Schedule	Purpose
1	Baseline	All enrolled mothers	Once at enrollment	- Collect baseline socio-demographic covariates (e.g. education, SES, occupation, household structure) - Collect household GPS coordinates - Identify health facilities potentially missed during census enumeration - Collect covariates relating to acceptability of phone (this section is administered to mothers in the phone group only)
2	Modified NFHS follow-up	All enrolled mothers	At each follow-up visit	- Assess for two-week prevalence of childhood illness and subsequent care seeking behavior - Collect information on timing of symptoms, type of care sought, and associated household costs (as applicable)
3	Special follow-up for phone group	Mothers enrolled in the phone group	At each follow-up visit	- Present mothers with a list of visits detected by GPS and rule out false positives - Link separate visits detected by GPS by individual episode
4	Health facility questionnaire	All enrolled health facilities	Once at enrollment	- Collect health facility GPS coordinates - Classify health facility by type - Collect covariates of potential relevance to care seeking (e.g. hours of operation, provider qualifications, services offered, payment schemes accepted)
5	Phone check-up	Mothers enrolled in the phone group	Once 3 days after enrollment, once 11-14 days after enrollment, and once at each follow-up visit	- Assess mothers' compliance with study procedures - Identify any problems mothers have encountered - Verify that optimal phone settings are in use

Table 3.1.5.1 Description of data included in each phone record

Variable name	Description
imei	International Mobile Equipment Identity (IMEI) number; this is a 15-digit number that will uniquely identify each phone; each study phone is assigned two unique IMEI numbers, one for each SIM slot
latitude	The latitude of the recorded GPS point; provided in decimal degrees (e.g. 18.6513)
longitude	The longitude of the recorded GPS point; provided in decimal degrees (e.g. 74.046898)
accuracy	The radius of 68% confidence around the recorded GPS coordinate; a higher value indicates lower accuracy; provided in meters
source	The phones use assisted-GPS, which allows for the detection of location through a combination of GPS satellites and cellular towers. The possible sources are: 1 = Fused 2 = GPS 3 = Cellular network
locationMode	The configuration of locations settings on the study phone at the time the record was saved: -1 = Unknown 0 = Location access disabled 1 = Device only 2 = Battery saving 3 = High accuracy
cached	The phone will occasionally recycle previously recorded (i.e. “cached”) coordinates; the exact circumstances when the phone uses the cached coordinates is not known, though it is thought to be a result of poor signal quality and the inability to fetch new coordinates
saveTime	The time when the GPS coordinate was recorded (milliseconds); this will be used in combination with the GPS coordinates to track the movement of the phone through time
sendTime	The time when the GPS coordinate was transferred to the study server (milliseconds)

Chapter 4: Determinants and Patterns of Care-seeking for Childhood Illness in Rural Pune District, India

Andrew Marsh MPH[*1, 2], Siddhivinayak Hirve PhD[2], Pallavi Lele MSc[2], Uddhavi Chavan MSc[2], Tathagata Bhattacharjee MBA[2, 3] Harish Nair PhD[4], Sanjay Juvekar PhD[2, 3†], Harry Campbell MD[4†]

[*] Corresponding author

[†] Joint last author with equal contribution

[1] Institute for International Programs, Johns Hopkins University Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD, USA

[2] KEM Hospital Research Centre, Sardar Moodliar Road, Rasta Peth, Pune 411011, India

[3] INDEPTH Network, 40 Mensah Wood Street, East Legon, Accra, Ghana

[4] Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, United Kingdom

Abstract

Background: An estimated 1.2 million children under five years of age die each year in India, with pneumonia and diarrhea among the leading causes. Increasing care-seeking is important to reduce mortality and morbidity from these causes. This paper explores the determinants and patterns of care-seeking for childhood illness in rural Pune district, India.

Methods: Mothers having at least one child < 5 years from the study area of the Vadu Health and Demographic Surveillance System were enrolled in a prospective cohort study. Household sociodemographic information was collected through a baseline questionnaire administered at enrollment. Participants were visited up to six times between July 2015 and February 2016 to collect information on recent childhood acute illness and associated care-seeking behavior. Multivariate logistic regression explored the associations between care-seeking and child, participant, and household characteristics.

Findings: We enrolled 743 mothers with 1,066 eligible children, completing 2,585 follow-up interviews (90% completion). Overall acute illness prevalence in children was 26% with care sought from a health facility during 71% of episodes. Multivariable logistic regression showed care-seeking was associated with the number of reported symptoms (adjusted odds ratio [AOR] = 2.4; 95% confidence interval [CI] 1.5-3.9; $p<0.01$) and household insurance coverage (AOR = 2.2; 95% CI 1.1-4.3; $p<0.05$). We observed an interaction between the associations of illness severity and maternal employment on care-seeking. Moderate-to-very severe illness was associated with increased care-seeking among both employed (AOR 5.0; 95% CI 2.2-11.1; $p<0.001$) and currently

unemployed mothers (AOR = 7.0; 95% CI 3.9-12.6; $p<0.001$). Maternal employment was associated with reduced care-seeking for non-severe illness (AOR 0.3, 95% CI 0.1-0.7; $p<0.001$), but not associated with care-seeking for moderate-to-very severe illness. Child sex was not associated with care-seeking.

Conclusions: This study demonstrates the importance of illness characteristics in determining facility-based care-seeking while also suggesting that maternal employment resulted in decreased care-seeking among non-severe illness episodes. The nature of the association between maternal employment and care-seeking is unclear and should be explored through additional studies. Similarly, the absence of male bias in care-seeking should be examined to assess for potential bias at other stages in the management of childhood illness.

Introduction

An estimated 5.9 million children under five years of age die each year globally with 1.2 million of these deaths occurring in India [1]. Acute respiratory infections (ARI) and diarrhea are among the leading causes of child mortality, yet many of these deaths may be avoided through scaling up coverage of effective interventions [2]. Care-seeking from a qualified health provider is a necessary step for the management of childhood illnesses, especially when appropriate treatment cannot be provided within the home [3]. Care-seeking for childhood illnesses in India has increased in the last decade, yet substantial gaps remain. An estimated 32% of children with diarrhea, 27% of children with fever, and 22% of children with symptoms of ARI were not brought for care in 2015-16, compared to 41%, 30%, and 31% of children with these illnesses in 2005-06 [61,74].

Data on care-seeking for childhood illness and associated determinants are frequently measured through large household surveys, such as the Demographic and Health Survey (DHS), implemented within India as the National Family and Health Survey (NFHS), Multiple Indicator Cluster Surveys, and Malaria Indicator Surveys [75-78]. A meta-analysis of 258 surveys from these sources administered between 2000 and 2013 identified common factors associated with care-seeking from a health facility, including child age, child sex, maternal education, urban residence, socioeconomic status (SES), and distance from the nearest health facility [9]. Care-seeking has been associated with illness severity, measured either according to caregiver perception or clinical criteria [10]. While medical professionals and policymakers may be most interested in clinically meaningful illness characteristics, evidence suggests that caregivers may be unable to accurately classify illnesses according to these criteria [79]. Furthermore, a multicountry study of care-seeking for childhood illness observed a strong association between the perception of illness and both illness symptoms and the decision to seek care [80].

Recent reviews of diarrheal disease and ARI in India noted variability in care-seeking practices by region, contrasting high care-seeking for childhood illness from predominantly western providers in the southern state of Kerala with the urban city of Lucknow in the northern India, where traditional healers were visited for 24% and 33% of episodes of neonatal ARI and diarrhea, respectively [62-65]. Care-seeking for childhood illness is higher in Maharashtra than in India as a whole with 89% of children with symptoms of ARI and 78% of children with diarrhea brought for care in Maharashtra compared to 78% and 68% of children with the same illnesses nationally [61]. Care is most frequently sought in the private sector, with wealthier households being more than twice as likely to seek care from a private provider [67]. A mixed-methods study of care-seeking for neonatal and childhood illnesses in rural Maharashtra noted that while private sector practitioners were the most common source of care for each condition, faith-based healers were frequently visited for selected neonatal danger signs (poor sucking, difficult breathing, and boils over body) and childhood measles [68].

Male bias has also been associated with care-seeking behavior with an analysis of national data showing increased delays in care-seeking for girls [81]. A study from rural Maharashtra found male bias across various stages of the care-seeking continuum, noting that caregivers of male children were more likely to seek care from a private practitioner, more likely to comply with referral, tended to spend more money on treatment, and were willing to travel further distance compared with caregivers of female children [69]. The most recent state-level data suggest no difference by child sex in facility-based care-seeking for childhood illness [61], though these data do not rule out male bias in provider choice, delays between onset and the decision to seek care, and household expenditures for care.

While large-scale household surveys collect data across a range of topics, their multipurpose approach restricts the detail they may provide on any single topic. For example, the NFHS questionnaire asks mothers of recently ill children whether care was sought, the facility types where care was sought, and which type of facility was accessed first if multiple types are indicated [61]. These data indicate care-seeking from a specific facility type and whether that type was the first contact with the health system but cannot describe the complete sequence of care-seeking events, including provider shopping behavior or repeated visits to the same provider type as has been explored elsewhere [17,82,83]. Additionally, studies evaluating the relationship between health facility access and care-seeking typically measure access according to participant-reported travel time or travel distance [20,22,23], though these may provide less precise measurements of access relative to more sophisticated GIS approaches.

As part of a study examining the validity of maternal recall of care-seeking for childhood illnesses in rural Pune district, India, we explored the determinants and patterns of care-seeking for diarrhea, fever, and cough among a cohort of children under five. Through an expanded questionnaire on care-seeking practices, we built on the approach of common household questionnaires to include data on sequential care-seeking actions, specific providers visited, and maternally-reported illness severity. This paper describes the results of data collected over six months of follow up with the specific objectives of identifying the factors associated with facility-based care-seeking and describing the sequential patterns of care-seeking for childhood illness.

Methods

Study Site

We conducted a prospective cohort study of care-seeking for childhood illness among mothers with young children in rural Pune district, India from May 2015 to February 2016. The study took place within the 22 villages of the Vadu Health and Demographic Surveillance System (HDSS), located 30 kilometers northeast of Pune city, Maharashtra state, India. The study area is served by seven public sector health facilities, including one rural hospital, one primary health center, and five sub-centers. The private sector includes 93 private hospitals and clinics and 68 pharmacies. Several non-facility public sector sources in the study area provide limited treatment for childhood illness, such as Anganwadi/Integrated Child Development Services (ICDS) centers and Accredited Social Health Activist (ASHA) workers.

Participant Enrollment and Follow-up

Study participants were mothers 15-49 years old with at least one child less than five years old, randomly sampled from the Vadu HDSS population register. Study recruitment occurred during field worker home visit. Consenting participants were randomly assigned to one of three study groups according to the objectives of the parent study. Participants assigned to the primary study group ("phone group") were provided with a GPS-enabled smartphone and followed up at six monthly visits. Participants in the longitudinal comparison group were similarly followed up at six monthly visits but not provided with smartphones. Participants in the cross-sectional comparison group were also not provided phones and were divided into six subgroups, each followed up once over the six-month follow-up period. Comparison groups were included to evaluate potential biases in reported care-seeking behavior among the phone group due to the presence of the smartphone or the repeated study contacts (see [84] and [85] for additional details). At each follow-up visit mothers were administered the NFHS module on childhood illness [61], which asks mothers whether their child experienced diarrhea, fever, or cough in the previous two weeks, whether any care was sought, and, if so, the sources from where care was sought. Mothers reporting a child with one or more of these symptoms were administered a supplementary questionnaire on symptom timing and the specific care-seeking actions taken in response, if any.

Measures

Our analysis focused on determinants of facility-based care-seeking, defined as any reported care-seeking from a public or private sector hospital or clinic. As in the NFHS definition, this included care provided by auxiliary nurse midwives at public sector sub-centers and excluded reported care-seeking from pharmacies, traditional healers, and shops. Additionally, we excluded care sought from Anganwadi/ICDS centers and ASHA workers, as these sources are limited in their capacity to provide comprehensive care for childhood illness.

Explanatory variables included characteristics of the illness, all of which were assessed at follow up, and characteristics of the child and household, which were assessed at baseline. Illness characteristics included maternally-reported severity (categorized as non-severe, moderately severe, or very severe as in [80]), number of reported symptoms (diarrhea, fever, or cough alone vs. multiple symptoms), and the presence of any danger signs (vomiting, difficulty eating, being unusually sleepy, or convulsions). Child variables included age at follow-up and sex. Household variables included maternal age, education, and employment status, number of children under-five in the household, religion of household head, SES, household structure (nuclear, extended), health insurance coverage, residence (urban, rural), and distance to the nearest health facility.

SES was determined according to the principal components analysis approach used by DHS [86,87]. Input variables included ownership of various assets (property, various durable goods, agricultural land, livestock, a bank/post office account, and health insurance or a health scheme), household building materials, drinking water source, toilet facility, and the presence of a servant or maid within the household. Distance to the nearest health facility was calculated as the shortest road-based distance using the inputs of participant and health facility locations coordinates and the current road network within the network analyst extension with ArcGIS 10 [88].

Statistical Analysis

We explored determinants and patterns of care-seeking in response to individual episodes of childhood illness, excluding records for children completing their fifth birthday prior to follow-up visit. We estimated the unadjusted and adjusted associations between individual predictors and care-seeking through univariate and multivariable logistic regression models, respectively. We accounted for correlation between repeated observations from the same child through the use of generalized estimating equations [89], specifying the binomial family, logit link, and an exchangeable correlation structure with robust estimation of standard errors. Variables were selected based on their previously demonstrated association with care-seeking. All variables were included in the final model except the presence of danger signs, which was excluded due to collinearity with illness severity and multiple reported symptoms. We also assessed for potential non-linear associations and effect modification, including an interaction term in the final model between illness severity and maternal employment status. Records with missing data (20%) were excluded from analysis according to list-wise deletion with the distribution of missing data by variable as follows: severity (14%), household structure (3%), maternal employment (2%), health insurance coverage (2%), and number of symptoms (<1%). Sensitivity analysis considered the influence of missing data through multiple imputation of missing values. Missing values were estimated according to univariate imputation models including care-seeking status and all other relevant predictors without missing data (e.g. study group, participant ID, child age, etc.). All statistical analyses were conducted in Stata 14 [90].

In addition to evaluating determinants of care-seeking, we also explored sequential patterns of care-seeking. Mothers who reported any care-seeking during an episode of childhood illness were asked to list each source from which care was sought, including the name, type, and number of days ago when care was sought. These data were linked to form sequences of care sought during an episode of childhood illness, starting with the first provider where care was sought and continuing with additional providers. Individual provider types were recoded into four major response categories to facilitate identification of common patterns: public facility, public non-facility, private facility and pharmacy. Care-seeking sequences were further stratified according to illness severity.

Ethical Approval

Written consent was provided by all participants prior to enrollment. The study protocol was approved by the ethics committees of the University of Edinburgh and K.E.M. Hospital Research Centre, Pune (Study ID No. 1415).

Results

Descriptive Analyses

Field workers enrolled 1,066 children ages 0-4 years from 743 households between May and June 2015 (Table 4.1). Mean child age at enrollment was 33 months (standard deviation [SD] 15) with 52% male children (n = 551). Mothers had a mean age at enrollment of 25.3 years (SD 3.3) with two thirds completing 10 or more years of schooling and a quarter reporting being currently employed. Mothers most frequently reported a single child under five (59%), followed by two (38%) and three children (3%), respectively.

Households were primarily Hindu religion (91%) with a similar proportion reporting extended and nuclear household structure and one fifth of households reporting some health insurance coverage. Two thirds of participant households were located in urban villages with 57% of participants less than one kilometer from the nearest health facility and 34% of participants between one and three kilometers from the nearest health facility (Figure 4.1). The nearest provider type was most often a private hospital or clinic.

Field workers completed 1,993 household visits between July 2015 and February 2016 (90% completion), administering the follow-up questionnaire to 2,761 children (Figure 4.2). Of these records, 176 (6%) were excluded as the child had completed his or her fifth birthday before follow-up. Mothers reported one or more symptom of childhood illness at 660 of the remaining 2,585 records (26%).

Completed care-seeking data were reported for 658 (99%) of these records, forming the sample for the remaining analyses. Fever was reported in 83% of instances of child illness, followed by cough (64%) and diarrhea (19%) (Figure 4.3, Supplementary Table 4.1). Childhood illnesses were frequently multi-symptomatic (60%) with combined fever and cough accounting for half of all reported illness episodes.

Table 4.2 presents care-seeking for childhood illnesses stratified by maternally reported illness severity. Care was sought from a health facility during 71% of illness episodes overall, increasing from 47% for non-severe illness to 88% for moderately severe illness and 100% for very severe illness. Care was primarily sought from the private sector across all severity strata with all episodes of very severe illness seeking care exclusively from the private sector.

Care-seeking Determinants

Table 4.3 presents descriptive characteristics of 467 cases where a child was brought for care and 191 cases where a child was not brought for care. Mothers of children brought for care were more likely to report moderate or very severe illness ($p<0.001$), more than one symptom ($p<0.001$), and the presence of one or more danger signs (p=0.005). Maternal employment was less common among cases for which a child was brought for care (p=0.028).

Table 4.4 presents the results of unadjusted and adjusted logistic regression analyses of care-seeking from a health facility. The adjusted analysis includes the interaction between illness severity and maternal employment status, strata specific estimates of which are presented in Table 4.5. Children with multiple symptoms were more than twice as likely to be brought for care as children with a single symptom (OR 2.4, 95% CI 1.5-3.9). Children with moderate and very severe illness were seven times more likely to seek care than children with non-severe illness in households where the mother was not employed (OR 7.0, 95% CI 3.9-12.6) and five times more likely to seek care in households where the mother was employed (OR 5.0, 95% CI 2.2-11.2). Maternal employment was associated with a significant decrease in care-seeking among cases of non-severe illness (OR 0.3, 95% CI 0.1-0.7) but not significantly associated with care-seeking among moderate and very severe cases. Children from a household where one or more member was covered by health insurance were more than twice as likely to be brought to a health facility for care (OR 2.2, 95% CI 1.1-4.3). Child age, child sex, maternal education, household socioeconomic status, urban residence, and proximity to a health facility were not associated with care seeking in either unadjusted or adjusted analyses. Sensitivity analysis applying multiple imputation detected no significant differences in coefficient estimates or confidence intervals for number of reported symptoms, perceived illness severity, maternal employment, or its interaction with perceived severity (Supplementary Tables 4.2-4.3, Supplementary Figure 4.1). A minor difference was observed in the estimated effect of health insurance coverage, resulting in this variable being marginally significant within the multiply imputed analysis (OR 1.8, 95% CI 1.0-3.3).

Care-seeking Patterns

A supplementary module on care-seeking steps was completed for 468 (91%) cases where care was sought from any source, including facility and non-facility sources. These data were linked to form sequences of care-seeking patterns, presented across all reported episodes and stratified by illness severity (Table 4.6). The most common care-seeking pattern overall and within each illness stratum was a single visit to a private provider, accounting for two thirds of care-seeking for non-severe cases and increasing to 79% among very severe cases. The second most common pattern overall and among moderate and very severe cases was care-seeking from two or more private sector providers, while the second most common pattern for non-severe illness was care-seeking exclusively from pharmacies and drugstores (18%). As illness severity increases, the proportion of care-seeking only from pharmacies or non-facility public sector sources decreases substantially, accounting for 21% of care-seeking for non-severe illness, 4% for moderately severe illness, and no care-seeking for very severe illness. We observed minimal crossover between sectors with only 1% of children (n=4) beginning in the private sector and less than 1% of children (n=2) beginning in the public sector brought for follow-up care in the opposite sector, respectively.

Discussion

Care-seeking for childhood illness in India has increased in the last decade both at the national level and within Maharashtra, where state-level data report care-seeking for diarrhea, fever, and symptoms of ARI during 78%, 85%, and 89% of episodes, respectively. Achieving further improvement in care-seeking within the state and nationally requires understanding the determinants and patterns of care-seeking for childhood illness. Through a population-based cohort of 1,066 children under-five in rural Pune district, India, we observed high levels of care-seeking overall with illness characteristics as the key determinants of care-seeking. Children whose illnesses were reported as moderate-to-very severe were seven times more likely to be brought for care than children whose illness was non-severe and children whose mothers reported they had more than one symptom were more than twice as likely to be brought for care.

One tenth of care-seeking involved self-medication through private pharmacies, typically as the only source of care though occasionally followed by facility-based care. While pharmacies may provide care for mild childhood illness (e.g. oral rehydration therapy for diarrhea), the quality of services they provide is often low and has been linked with medication misuse [91-93]. A review of pharmacy performance in Asian countries observed poor adherence to guidelines for the treatment of diarrhea with pharmacists recommending oral rehydration solution less than half of the time while frequently providing antibiotics and other unnecessary medicines [94]. Within our study, the pattern of pharmacy-based care varied with perceived illness severity. Nearly 20% of children with non-severe illness were brought to pharmacies and drugstores for care compared to 6% of children with moderate illness and no children with very severe illness. Decreased care-seeking from pharmacies among more severe illness is consistent with a recent review of medication misuse in India, reporting that self-medication was preferred among mild illness as a strategy to avoid the cost of doctor consultation or diagnostic tests [93]. Similarly, mild perceived severity was associated with the decision to seek care directly from a pharmacy among adults with symptoms of ARI in Bangladesh [95].

No association was observed between child sex and care-seeking for childhood illness. This is consistent with the most recent care-seeking estimates for Maharashtra, where similar proportions of facility-based care-seeking were observed by child sex for diarrhea, fever, and symptoms of ARI [61]. Similar care-seeking by child sex has also been reported elsewhere in India [96]. While sex did not appear to be associated with the decision to seek care, this does not exclude potential bias at other stages of the management of childhood illness. Previous findings from the study area identified male preference in utilization of private providers, distance traveled for care, money spent, and compliance with referral [69]. A study in Uttar Pradesh found no association between child sex and care-seeking for neonatal illness but noted both delays in illness recognition and significantly less money spent on girls [97]. Exploring sex bias across all stages in the management of childhood illness was beyond the scope of the study, which restricted its focus to facility-based care-seeking. Further research is needed to assess for bias across the other steps in the pathway from illness recognition to type of care sought and compliance with referral.

Children with non-severe illness were 70% less likely to be brought for care if their mother was currently employed, while maternal employment was not significantly associated with care-seeking for moderate-to-very severe illness. Previous studies exploring the role of maternal employment in health utilization have found mixed results. An analysis of national data in India linked improved child immunization status and care-seeking practices with high maternal autonomy, measured as financial access, freedom of movement, and decision-making power [98].

Sometimes used as a proxy for maternal autonomy, maternal employment was included as a separate covariate in the analysis and was found to be associated with both poorer immunization status and decreased care-seeking for childhood illness [98]. In contrast, maternal employment was associated with increased care-seeking for obstetric complications among women in Bangladesh [99]. Within our study context, the interaction between employment and illness severity may indicate that the opportunity cost of seeking care for employed mothers is sufficient to result in lower care-seeking for non-severe illness but insufficient to deter mothers from seeking care for more severe illness.

Maternal education, access to care, and SES have previously been found to be associated with care-seeking for childhood illness but did not appear significant within our study. This should be interpreted alongside the baseline distributions of these characteristics within the study population. Mothers in our study area are highly educated with three quarters completing 10 or more years of schooling, compared to one quarter of similarly aged women in Maharashtra [61]. Household access to care was also high with more than half of households located within one kilometer of health facility and most remaining households located less than three kilometers from a health facility. While not explicit measurements of household SES, high levels of education and access to care would suggest higher than average SES within the study population relative to state and national estimates. As underlying SES of the study population increases, the increment between calculated SES quintiles becomes less meaningful. For example, the lowest and highest quintiles in our sample may still represent the poorest and wealthiest households, respectively, but if the poorest are still relatively wealthy then we would not expect to see much difference between groups with regard to care-seeking or other variables typically associated with SES. Our null findings with regard to these variables may therefore rule out the potential for a large effect size but may be consistent with a smaller one.

This study includes limitations. First, data on childhood illness and related care-seeking behavior during the previous two weeks were collected during participant interviews and may be subject to a combination of recall and social desirability bias. A recent study of the validity of maternally-reported care-seeking for childhood illness found that the indicator had both high sensitivity and specificity, resulting in only a minor overestimation of care-seeking relative to true levels [59]. Previous experience within this study area noted the potential for underreporting of sensitive health behaviors (e.g. abortion services), especially when solicited through a survey-based approach [100], though the same degree of bias is unlikely when asking about less sensitive behaviors. Second, we classified illness severity according to mothers' perception rather than through the presence of clinically-defined symptoms. This decision was taken in part because recent evidence suggests mothers have low discriminative power when differentiating between clinically meaningful presentations of similar symptoms [79,101]. The perception of severity has been associated with actual severity, though culturally-specific interpretations of signs and symptoms may also lead to concern [80]. We also expect that illness perception contributes more directly to a mother's decision to seek care than do the underlying clinical symptoms. Third, the high levels of sociodemographic indicators (e.g. mother's education, access to care) observed within our study population relative to recent state-level estimates limit the generalizability of our findings to the state of Maharashtra and similar contexts [61]. The study area represents a community in transition, historically based in agriculture but subject to the urbanizing influence of the nearby metropolitan area. As large urban centers continue to expand throughout India and elsewhere, the number of such communities will continue to grow. Finally, one fifth of illness records included at least one variable with a missing response with illness severity missing most frequently. Excluding these records from analysis may have reduced statistical efficiency and yielded null results where a significant association might otherwise be detected. Similarly, excluding records with missing values may have resulted in biased estimates if these data were systematically different from those that were not missing. Our sensitivity analysis comparing completed cases alone with the multiply imputed dataset yielded similar results, suggesting no significant bias in our results due to missing data.

Conclusion

This study demonstrates the importance of maternal perception of illness severity in determining facility-based care-seeking for childhood illness in rural Pune district, India. A further association between maternal employment and decreased care-seeking was noted in non-severe cases, though not in moderate-to-very severe cases. Additional research is required to determine whether male bias exists at different stages in the management of childhood illness, such as illness recognition and resources allocated to treatment. As care-seeking for childhood illness continues to rise in India, there is an urgent need to develop indicators to assess the content of care, especially with regard to the appropriateness of treatments for the illness being treated [102].

Acknowledgements

Vadu HDSS is a member of the INDEPTH Network. We are grateful to the members of the Improving Coverage Measurement Core Group for their guidance in developing this study. We also thank Drs. Robert Black, Joanne Katz, Krishna Rao and Melinda Munos of the Johns Hopkins School of Public Health for their comments on the study protocol and this paper. We acknowledge Sandeep Bhujbal for developing and maintaining the data management systems for the project. Lastly, we wish to thank the study participants and the project staff at the Vadu HDSS, without whom this study would not have been possible.

Source of Funding

The study was funded by a prime award by the Bill and Melinda Gates Foundation (grant no. OPP1084442) made to the Johns Hopkins University of which University of Edinburgh, Edinburgh, UK and KEM Hospital Research Centre, Pune, India were sub-grantees.

Author Contribution

All authors conceived of the study and protocol. AM, PL, UC, and SJ conducted data collection. AM analyzed the data and wrote the paper. SH, HN, SJ, and HC provided guidance on the analysis and interpretation of results. All authors read and agree with the manuscript and conclusions.

Competing Interest

The authors declare no competing interest

References

1 Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet (London, England). 2016;388:3027-35.

2 Jones G, Steketee RW, Black RE, Bhutta ZA, Morris SS. How many child deaths can we prevent this year? The Lancet. 2003;362:65-71.

3 Waldman RJ, Bartlett AV, Campbell CC, Steketee RW. Overcoming remaining barriers: the pathway to survival. Arlington, VA: BASICS Project, 1996.

9 Bennett A, Eisele T, Keating J, Yukich J. Global trends in care seeking and access to diagnosis and treatment of childhood illnesses. Rockville, Maryland, USA: ICF International, 2015.

10 Geldsetzer P, Williams TC, Kirolos A, Mitchell S, Ratcliffe LA, Kohli-Lynch MK, et al. The recognition of and care seeking behaviour for childhood illness in developing countries: a systematic review. PLoS One. 2014;9:e93427.

17 Baume C, Helitzer D, Kachur SP. Patterns of care for childhood malaria in Zambia. Soc Sci Med. 2000;51:1491-503.

20 Glik DC, Ward WB, Gordon A, Haba F. Malaria treatment practices among mothers in Guinea. Journal of health and social behavior. 1989;30:421-35.

22 Das A, Ravindran TS. Factors affecting treatment-seeking for febrile illness in a malaria endemic block in Boudh district, Orissa, India: policy implications for malaria control. Malar J. 2010;9:377.

23 Chaturvedi HK, Mahanta J, Pandey A. Treatment-seeking for febrile illness in north-east India: an epidemiological study in the malaria endemic zone. Malar J. 2009;8:301.

59 Carter ED, Ndhlovu M, Munos M, Nkhama E, Katz J, Eisele TP. Validity of maternal report of care-seeking for childhood illness. J Glob Health. 2018;8:010602.

61 International Institute for Population Sciences (IIPS) and ICF. National Family Health Survey (NFHS-4), 2015-16: India. Mumbai: IIPS, 2017.

62 Mathew JL, Patwari AK, Gupta P, Shah D, Gera T, Gogia S, et al. Acute respiratory infection and pneumonia in India: a systematic review of literature for advocacy and action: UNICEF-PHFI series on newborn and child health, India. Indian pediatrics. 2011;48:191-218.

63 Shah D, Choudhury P, Gupta P, Mathew JL, Gera T, Gogia S, et al. Promoting appropriate management of diarrhea: a systematic review of literature for advocacy and action: UNICEF-PHFI series on newborn and child health, India. Indian pediatrics. 2012;49:627-49.

64 Pillai RK, Williams SV, Glick HA, Polsky D, Berlin JA, Lowe RA. Factors affecting decisions to seek treatment for sick children in Kerala, India. Soc Sci Med. 2003;57:783-90.

65 Awasthi S, Srivastava NM, Pant S. Symptom-specific care-seeking behavior for sick neonates among urban poor in Lucknow, Northern India. J Perinatol. 2008;28 Suppl 2:S69-75.

67 Sreeramareddy CT, Sathyanarayana TN, Kumar HN. Utilization of health care services for childhood morbidity and associated factors in India: a national cross-sectional household survey. PLoS One. 2012;7:e51904.

68 Dongre AR, Deshmukh PR, Garg BS. Childhood morbidity, household practices and health care seeking for sick children in a tribal district of Maharashtra, India. Indian J Med Sci. 2010;64:7-16.
69 Ganatra B, Hirve S. Male bias in health care utilization for under-fives in a rural community in western India. Bull World Health Organ. 1994;72:101-4.
74 International Institute for Population Sciences (IIPS) and Macro International. National Family Health Survey (NFHS-3), 2005-06: India: Volume 1. Mumbai: IIPS: 2007.
75 MEASURE DHS. Demographic and Health Surveys. Available: http://www.measuredhs.com/. Accessed: May 16, 2018.
76 International Institute for Population Sciences (IIPS). National Family Health Survey. Available: http://rchiips.org/nfhs/. Accessed: May 17, 2018.
77 UNICEF. Multiple Indicator Cluster Survey (MICS). Available: http://www.unicef.org/statistics/index_24302.html Accessed: May 16, 2018.
78 Roll Back Malaria. Malaria Indicator Surveys. Available: http://www.malariasurveys.org. Accessed: May 16, 2018.
79 Hazir T, Begum K, El Arifeen S, Khan AM, Huque MH, Kazmi N, et al. Measuring coverage in MNCH: a prospective validation study in Pakistan and Bangladesh on measuring correct treatment of childhood pneumonia. PLoS Med. 2013;10:e1001422.
80 Yoder PS, Hornik RC. Symptoms and perceived severity of illness as predictive of treatment for diarrhea in six Asian and African sites. Social Science & Medicine. 1996;43:429-39.
81 Malhotra N, Upadhyay RP. Why are there delays in seeking treatment for childhood diarrhoea in India? Acta Paediatr. 2013;102:e413-8.
82 Akin JS, Hutchinson P. Health-care facility choice and the phenomenon of bypassing. Health Policy Plan. 1999;14:135-51.
83 Ryan GW. What do sequential behavioral patterns suggest about the medical decision-making process?: modeling home case management of acute illnesses in a rural Cameroonian village. Soc Sci Med. 1998;46:209-25.
84 Nair H, Williams L, Marsh A, Lele P, Bhattacharjee T, Chavan U, et al. Assessing the impact of mobile phones and repeated surveys on healthcare seeking behaviour for common childhood illnesses in rural India. 2018.
85 Hirve S, Marsh A, Lele P, Chavan U, Bhattacharjee T, Nair H, et al. Concordance between GPS-based smartphone app for continuous location tracking and mother's recall of care-seeking for child illness in India. 2018.
86 Filmer D, Pritchett LH. Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India. Demography. 2001;38:115-32.
87 Rutstein SO, Johnson K. The DHS wealth index. Calverton, Maryland, USA: ORC Macro, 2004.
88 ESRI. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute; 2011.
89 Liang KY, Zeger SL. Regression analysis for correlated data. Annual review of public health. 1993;14:43-68.
90 StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP; 2015.

91 Smith F. The quality of private pharmacy services in low and middle-income countries: a systematic review. Pharm World Sci. 2009;31:351-61.
92 Chandy SJ. Consequences of irrational use of antibiotics. Indian journal of medical ethics. 2008;5:174-5.
93 Porter G, Grills N. Medication misuse in India: a major public health issue in India. Journal of public health (Oxford, England). 2016;38:e150-7.
94 Miller R, Goodman C. Performance of retail pharmacies in low- and middle-income Asian settings: a systematic review. Health Policy Plan. 2016;31:940-53.
95 Chowdhury F, Sturm-Ramirez K, Mamun AA, Iuliano AD, Bhuiyan MU, Chisti MJ, et al. Factors driving customers to seek health care from pharmacies for acute respiratory illness and treatment recommendations from drug sellers in Dhaka city, Bangladesh. Patient Prefer Adherence. 2017;11:479-86.
96 Manna B, Nasrin D, Kanungo S, Roy S, Ramamurthy T, Kotloff KL, et al. Determinants of health care seeking for diarrheal illness in young children in urban slums of Kolkata, India. Am J Trop Med Hyg. 2013;89:56-61.
97 Willis JR, Kumar V, Mohanty S, Singh P, Singh V, Baqui AH, et al. Gender differences in perception and care-seeking for illness of newborns in rural Uttar Pradesh, India. Journal of health, population, and nutrition. 2009;27:62-71.
98 Malhotra C, Malhotra R, Ostbye T, Subramanian SV. Maternal autonomy and child health care utilization in India: results from the National Family Health Survey. Asia Pac J Public Health. 2014;26:401-13.
99 Sikder SS, Labrique AB, Craig IM, Wakil MA, Shamim AA, Ali H, et al. Patterns and determinants of care seeking for obstetric complications in rural northwest Bangladesh: analysis from a prospective cohort study. BMC Health Serv Res. 2015;15:166.
100 Ganatra B, Hirve S. A community-based study on induced abortions. Issues Med Ethics. 2001;9:7-8.
101 Choi Y, El Arifeen S, Mannan I, Rahman SM, Bari S, Darmstadt GL, et al. Can mothers recognize neonatal illness correctly? Comparison of maternal report and assessment by community health workers in rural Bangladesh. Trop Med Int Health. 2010;15:743-53.
102 Marchant T, Bryce J, Victora C, Moran AC, Claeson M, Requejo J, et al. Improved measurement for mothers, newborns and children in the era of the Sustainable Development Goals. J Glob Health. 2016;6:010506.

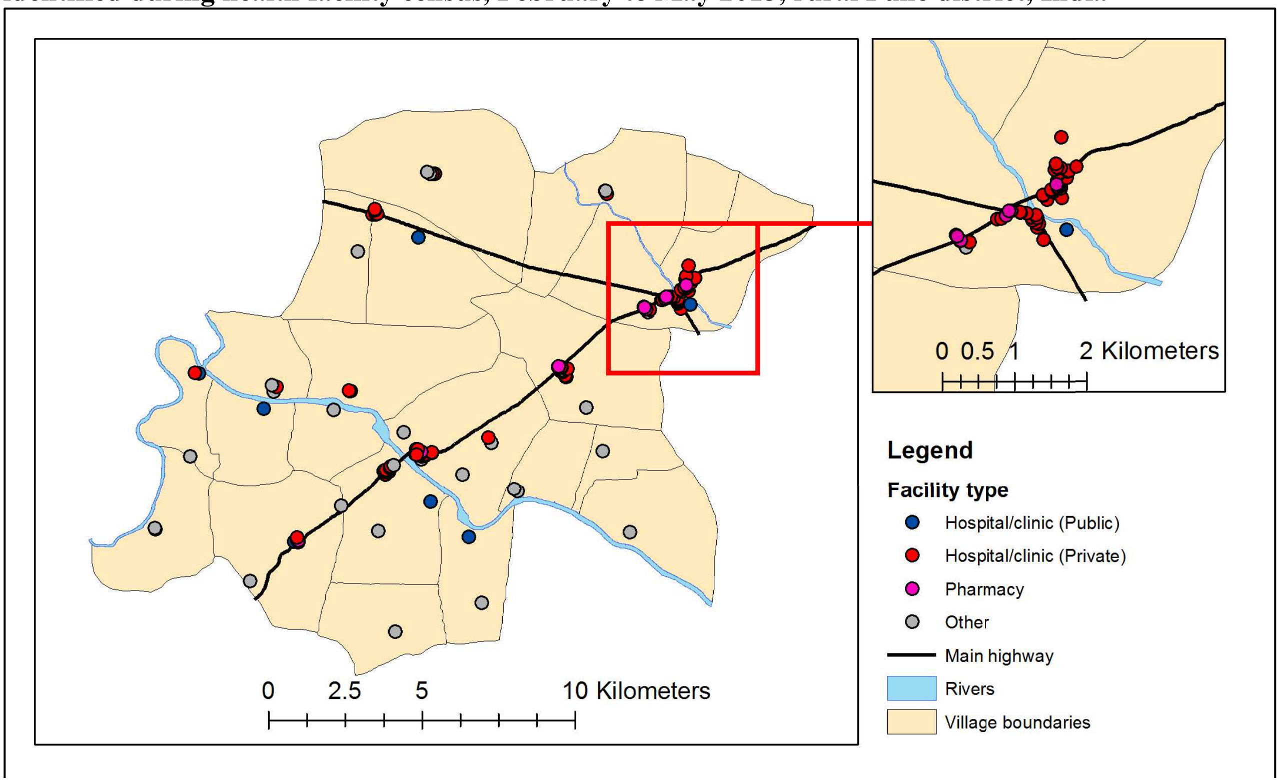
Tables and Figures

Table 4.1. Baseline child and household characteristics, May-June 2015, rural Pune district, India

Characteristic	N (%)
Child Characteristics (n=1066)	
Child age, months, mean (SD)	33 (15)
Child sex	
Male	551 (52)
Female	515 (48)
Household characteristics (n=759)	
Maternal age, years, mean (SD)	25.3 (3.3)
Maternal education, years completed	
0-7	111 (15)
8-9	114 (15)
10-11	212 (29)
12+	306 (41)
Mother currently employed	206 (28)
Number of children < 5 years per household	
One	442 (59)
Two	279 (38)
Three	22 (3)
Religion of household head	
Hindu	678 (91)
Muslim	23 (3)
Buddhist or Neo-Buddhist	34 (5)
Other	8 (1)
Family structure	
Extended	365 (51)
Nuclear	348 (49)
Insurance ownership	135 (19)
Urban residence	506 (68)
Distance to nearest health facility (public or private), km	
<1	426 (57)
1-3	251 (34)
>3	66 (9)
Nearest facility by sector	
Public	103 (14)
Private	640 (86)

Km = kilometer

Figure 4.1. Study area of the Vadu Health and Demographic Surveillance System with all facilities identified during health facility census, February to May 2015, rural Pune district, India



Note: Public sector hospitals and clinics include the rural hospital (1), primary health center (1), and sub-centers (7). Private sector hospitals and clinics include private hospitals (49), private clinics (43), and the non-governmental organization/trust hospital (1). Pharmacies include only pharmacies/drugstores (68). Other facilities include Anganwadi/ICDS centers (24) and shops (4).

Figure 4.2. Flow diagram of participant enrollment and follow-up

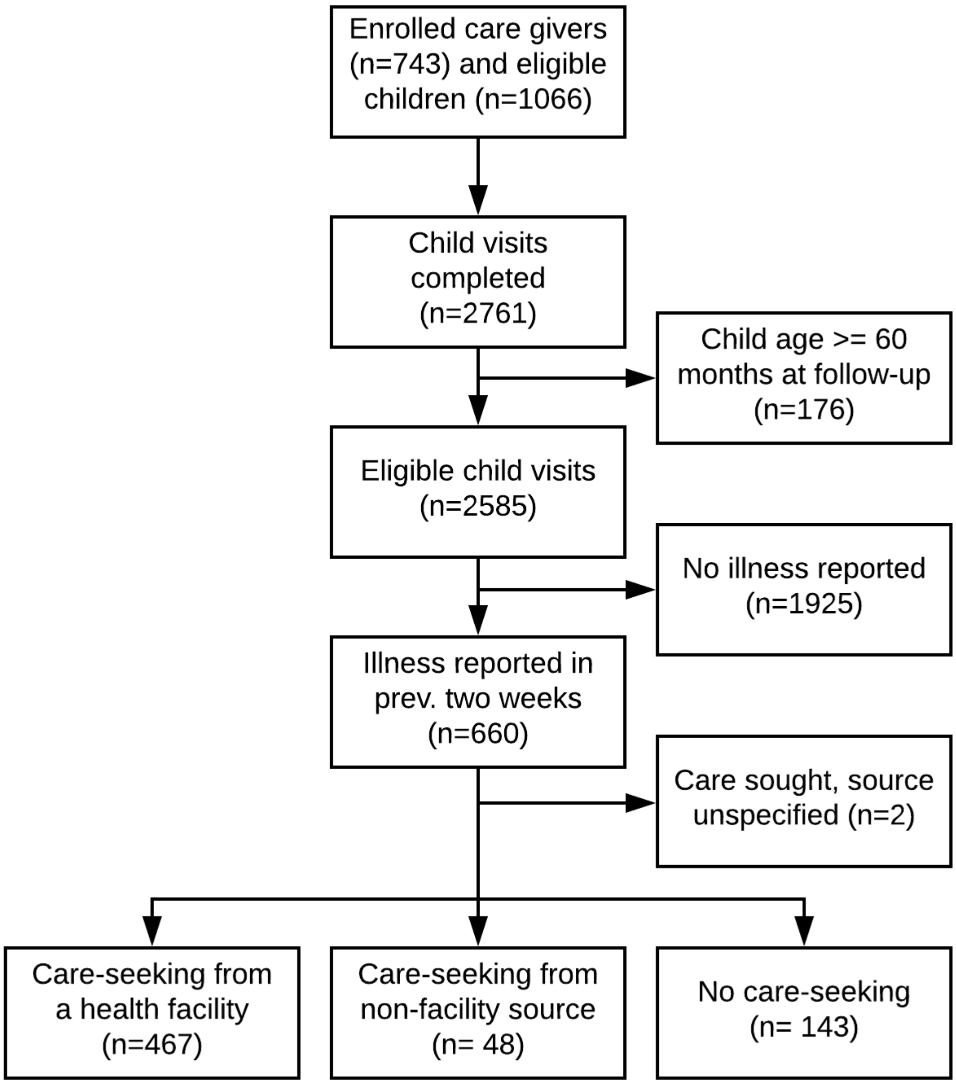


Figure 4.3. Illness profile of 658 children reporting one or more symptoms of childhood illness in the previous two weeks between July 2015 to February 2016 in rural Pune district, India

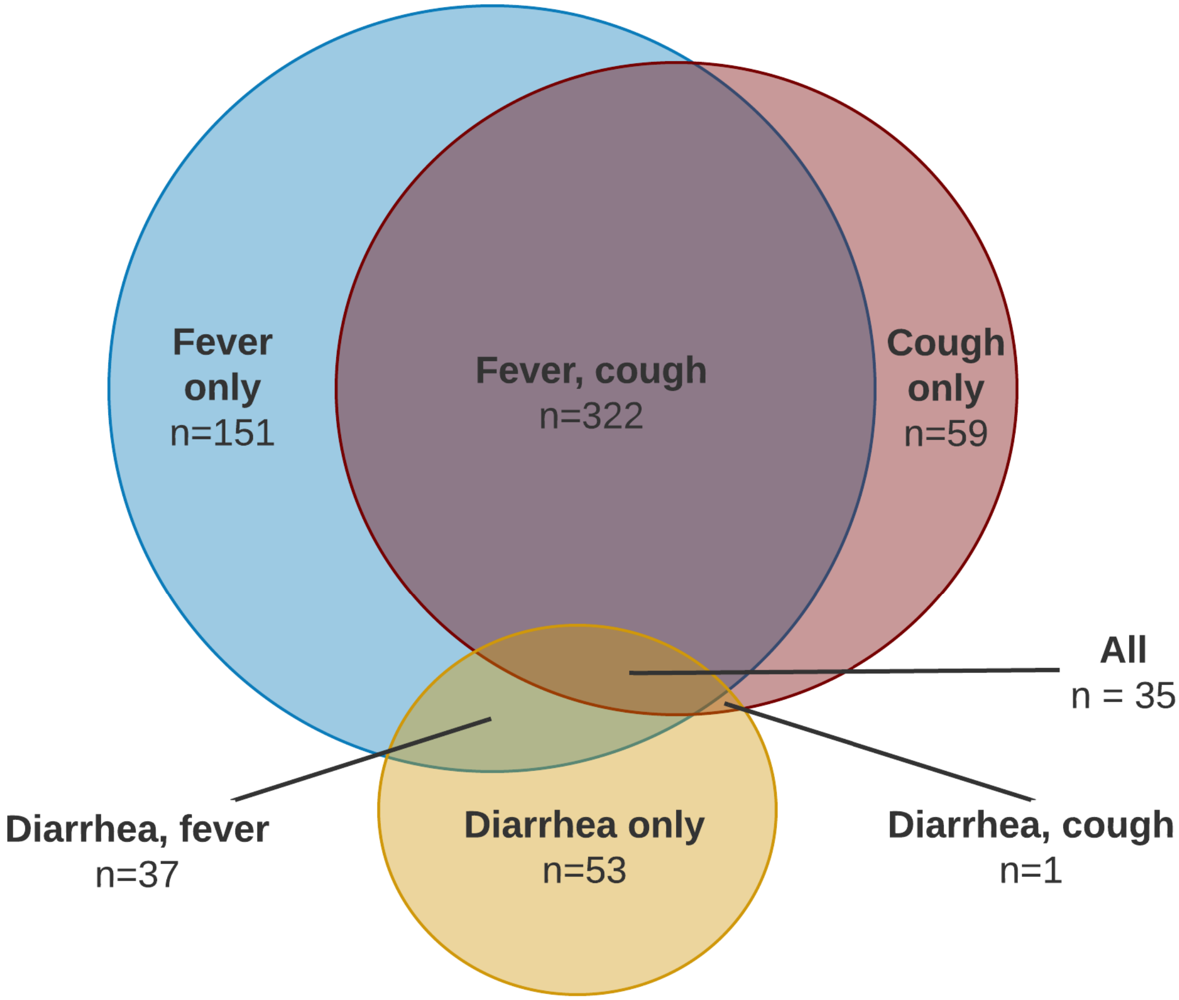


Table 4.2. Care-seeking by illness severity for 658 childhood illness episodes reported between July 2015 and February 2016, rural Pune district, India

Care seeking	All cases N (%)	Non-severe N (%)	Moderately severe N (%)	Very severe N (%)	Severity missing N (%)
No care sought	143 (22)	96 (40)	22 (7)	0 (0)	25 (27)
Care sought from any source	515 (78)	142 (60)	272 (93)	32 (100)	69 (73)
Care sought, health facility[1]	467 (71)	113 (47)	258 (88)	32 (100)	64 (68)
Public sector facility	35 (5)	7 (3)	22 (7)	0 (0)	6 (6)
Private sector facility	438 (67)	106 (45)	242 (82)	32 (100)	58 (62)
Care sought, non-facility source	48 (7)	29 (13)	14 (5)	0 (0)	5 (5)
Total	658 (100)	238 (100)	294 (100)	32 (100)	94 (100)

[1] Sum of care sought from public and private sector facilities may exceed care sought from health facility due to multiple reported care-seeking events per illness episode

Note: Public sector facility sources include: rural hospital, primary health center, and sub-centre/auxiliary nurse midwife. Private health facility sources include: private hospital, private doctor/clinic, and non-governmental organization or trust hospital/clinic

Table 4.3. Descriptive characteristics of participants seeking care and not seeking care for childhood illness from July 2015 to February 2016, rural Pune district, India

Characteristic	Care sought (N=467) N (%)	Care not sought (N=191) N (%)	P-value
Child characteristics			
Child age, months, mean (SD)	33.4 (14.5)	32.9 (14)	0.66
Child sex			
Male	234 (50)	95 (50)	0.93
Female	233 (50)	96 (50)	
Perceived severity			
Very severe	32 (8)	0 (0)	<0.001
Moderately severe	258 (64)	36 (22)	
Non-severe	113 (28)	125 (78)	
Number of reported symptoms			
Single	148 (32)	113 (60)	<0.001
Multiple	318 (68)	77 (40)	
Reported danger signs			
No	402 (89)	168 (96)	0.005
Yes	51 (11)	7 (4)	
Maternal characteristics			
Maternal education, completed years			
0-7	65 (14)	34 (18)	0.25
8-9	71 (15)	19 (10)	
10-11	142 (30)	60 (31)	
12+	189 (41)	78 (41)	
Maternal employment status			
Not currently employed	358 (78)	128 (70)	0.03
Currently employed	102 (22)	56 (30)	
Household characteristics			
Number of CU5 in household			
Single	222 (48)	79 (41)	0.15
Multiple	245 (52)	112 (59)	
Household structure			
Extended	244 (54)	90 (49)	0.33
Nuclear	210 (46)	92 (51)	
Religion of head of household			
Hindu	417 (89)	165 (86)	0.29
Other	50 (11)	26 (14)	
Wealth quintile			
Quintile 1 (lowest)	92 (20)	46 (24)	0.62
Quintile 2	69 (15)	26 (14)	
Quintile 3	102 (22)	38 (20)	
Quintile 4	110 (24)	49 (26)	
Quintile 5 (highest)	94 (20)	32 (17)	
Health insurance coverage			
No	355 (77)	155 (83)	0.11
Yes	105 (22)	32 (17)	
Residence			
Rural	162 (35)	64 (34)	0.77
Urban	305 (65)	127 (66)	
Distance to nearest health facility			
<1 km	280 (60)	115 (60)	0.68
1-3 km	146 (31)	63 (33)	
>3km	41 (9)	13 (7)	

Note: Difference in child's age at follow-up calculated using Student's T-Test. Differences for all other variables calculated using Pearson's Chi-squared test. Tests are not clustered by participant.

Table 4.4. Univariate and multivariable logistic regression results of care-seeking for childhood illness between July 2015 to February 2016, rural Pune district, India

Characteristic	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Multiple symptoms reported	3.2 (2.2 - 4.6)**	2.4 (1.5 - 3.9)**
Presence of danger signs	3.1 (1.4 - 6.9)**	See note
Illness perceived as moderate-to-very severe	8.8 (5.5 - 13.9)**	7.0 (3.9 - 12.6)**
Mother currently employed	0.6 (0.4 - 0.9)*	0.3 (0.1 - 0.7)**
Interaction: severity x maternal employment	N/A	2.3 (0.8 - 6.8)
Maternal education, completed years		
0-7	REF	REF
8-9	1.7 (0.9 - 3.3)	1.4 (0.5 - 3.7)
10-11	1.2 (0.7 - 2.1)	1.0 (0.5 - 2.2)
12+	1.2 (0.7 - 2.0)	0.8 (0.4 - 1.8)
Child age, months	1.0 (1.0 - 1.0)	1.0 (1.0 - 1.0)
Child being female	1.0 (0.7 - 1.5)	1.0 (0.6 - 1.7)
Other children under-five in household	0.8 (0.6 - 1.2)	0.8 (0.5 - 1.2)
Household structure		
Extended	REF	REF
Nuclear	0.8 (0.6 - 1.2)	0.7 (0.4 - 1.4)
Religion of head of household		
Hindu	REF	REF
Other	0.8 (0.5 - 1.3)	0.8 (0.4 - 1.6)
Wealth quintile		
Quintile 1 (lowest)	REF	REF
Quintile 2	1.3 (0.7 - 2.4)	1.2 (0.5 - 2.8)
Quintile 3	1.3 (0.7 - 2.3)	1.4 (0.6 - 3.1)
Quintile 4	1.1 (0.6 - 1.8)	0.9 (0.4 - 2.1)
Quintile 5 (highest)	1.4 (0.8 - 2.5)	1.8 (0.6 - 4.8)
Covered by health scheme or health insurance	1.4 (0.9 - 2.3)	2.2 (1.1 - 4.3)*
Urban residence	1.0 (0.7 - 1.5)	1.2 (0.6 - 2.3)
Distance to nearest health facility		
<1 km	REF	REF
1-3 km	1.0 (0.6 - 1.4)	1.1 (0.6 - 2)
>3km	1.4 (0.7 - 2.7)	0.9 (0.3 - 2.5)

† $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; OR = odds ratio; CI = confidence interval; REF = reference group; km = kilometer

Note: Presence of danger signs excluded from final model due to collinearity with other covariates. Confidence intervals estimated using robust standard errors.

Table 4.5. Interaction between illness severity and maternal employment on care-seeking for childhood illness between July 2015 and February 2016, rural Pune district, India

	Illness severity		
	Non-severe OR (95% CI)	Moderate-to-very severe OR (95% CI)	*Effect of severity within employment strata* OR (95% CI)
Maternal employment			
Not currently employed	1.0 (Reference)	7.0 (3.9 - 12.6)**	7.0 (3.9 - 12.6)**
Currently employed	0.3 (0.1 - 0.7)**	5.0 (2.3 – 11.2)**	16.4 (6.7 - 40.2)**
Effect of employment status within severity strata	0.3 (0.1 - 0.7)**	0.7 (0.3 - 1.6)	

† $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; OR = odds ratio; CI = confidence interval

Measure of interaction on multiplicative scale: ratio of ORs (95% CI) = 2.3 (0.8 - 6.8)

Note: ORs are adjusted for number of reported symptoms, maternal education, child age, child sex, household structure, religion of head of household, household SES, health insurance coverage, urban residence, and distance to nearest health facility. Confidence intervals estimated using robust standard errors.

Table 4.6. Sequential care-seeking patterns by illness severity for 468 illness episodes reporting any care-seeking and completing supplemental questionnaire between July 2015 and February 2016, rural Pune district, India

Care seeking sequence	All cases N (%)	Non-severe N (%)	Moderately severe N (%)	Very severe N (%)	Severity missing N (%)
Pvt. Fac. only	334 (71)	89 (67)	186 (73)	22 (79)	37 (70)
Pvt. Fac.-->Pvt. Fac.[1]	51 (11)	8 (6)	30 (12)	6 (21)	7 (13)
Pharm.[2, 3] only	38 (8)	24 (18)	10 (4)	-	4 (8)
Pub. Fac. only	29 (6)	7 (5)	17 (7)	-	5 (9)
Pharm.-->Pvt. Fac.	5 (1)	1 (1)	4 (2)	-	-
Other Pub.[3] only	5 (1)	4 (3)	1 (<1)	-	-
Pvt. Fac.-->Pub. Fac.	4 (1)	-	4 (2)	-	-
Pub. Fac.-->Pvt. Fac.	2 (<1)	-	2 (1)	-	-
Pharm.-->Pvt. Fac.-->Pharm.	1 (<1)	-	1 (<1)	-	-
Total	468 (100)	133 (100)	254 (100)	28 (100)	53 (100)

Pvt. Fac. = Private sector facility, including private hospital, private doctor/clinic, and non-governmental organization or trust hospital/clinic. **Pharm.** = Pharmacy/drugstore. **Pub. Fac.** = Public sector facility, including rural hospital, primary health center, and sub-centre/auxiliary nurse midwife. **Other Pub.** = Non-facility public sector source, including Anganwadi/Integrated Child Development Center and Accredited Social Health Activist workers.

[1] Includes eight records with three or more care-seeking events exclusively from private facilities

[2] Includes two records with two care-seeking events exclusively at pharmacies

[3] Care sought only from these non-facility sources is excluded when calculating care seeking from a health facility

Supplementary Tables and Figures

Supplementary Table 4.1 Illness characteristics of 658 childhood illness episodes reported between July 2015 and February 2016, rural Pune district, India

Characteristic	N (%)
Single symptom	263 (40)
Diarrhea only	53 (8)
Fever only	151 (23)
Cough only	59 (9)
Suspected ARI	3 (0)
Multiple symptoms	395 (60)
Diarrhea, fever	37 (6)
Diarrhea, cough	1 (0)
Fever, cough	322 (49)
Fever, Suspected ARI	30 (5)
Diarrhea, fever, cough	35 (5)
Diarrhea, fever, Suspected ARI	7 (1)
Presence of danger signs	63 (10)
Vomiting	41 (7)
Difficulty eating	12 (2)
Child unusually sleepy/unconscious	5 (1)
Lower chest in-drawing	5 (1)
Convulsions	0 (0)
Perceived severity	
Non-severe	32 (5)
Moderately severe	294 (47)
Very severe	238 (38)
Missing	94 (14)
Status at follow-up interview	
Resolved	403 (64)
Not resolved	225 (36)
Symptom onset, mean days prior to interview (SD)	
Resolved episodes	9 (4)
Unresolved episodes	5 (4)
Illness duration[1], mean days (SD)	4 (2)

[1]Illness duration calculated for resolved episodes only

Supplementary Table 4.2. Multiply imputed univariate and multivariable logistic regression results of care-seeking for childhood illness between July 2015 to February 2016, rural Pune district, India

Characteristic	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Multiple symptoms reported	3.2 (2.2 - 4.5)**	2.5 (1.7 - 3.9)**
Presence of danger signs	3.1 (1.4 - 7.1)**	See note
Illness perceived as moderate-to-very severe	8.6 (5.5 - 13.4)**	7.3 (4.2 - 12.9)**
Mother currently employed	0.6 (0.4 - 0.9)*	0.3 (0.2 - 0.7)**
Interaction: severity x maternal employment	N/A	2.0 (0.7 - 5.6)
Maternal education, completed years		
0-7	REF	REF
8-9	1.7 (0.9 - 3.3)	1.7 (0.7 - 4.1)
10-11	1.2 (0.7 - 2.1)	1.1 (0.6 - 2.3)
12+	1.2 (0.7 - 2)	1.0 (0.5 - 2.1)
Child age, months	1.0 (1.0 – 1.0)	1.0 (1.0 – 1.0)
Child being female	1.0 (0.7 - 1.5)	1.1 (0.7 - 1.7)
Other children under-five in household	0.8 (0.6 - 1.2)	0.8 (0.5 - 1.2)
Household structure		
Extended	REF	REF
Nuclear	0.8 (0.6 - 1.2)	0.7 (0.4 - 1.3)
Religion of head of household		
Hindu	REF	REF
Other	0.8 (0.5 - 1.3)	0.8 (0.4 - 1.5)
Wealth quintile		
Quintile 1 (lowest)	REF	REF
Quintile 2	1.3 (0.7 - 2.5)	1.2 (0.6 - 2.4)
Quintile 3	1.3 (0.7 - 2.3)	1.6 (0.8 - 3.3)
Quintile 4	1.1 (0.6 - 1.9)	1.0 (0.4 – 2.0)
Quintile 5 (highest)	1.4 (0.8 - 2.5)	1.7 (0.7 - 4.2)
Covered by health scheme or health insurance	1.4 (0.9 - 2.3)	1.8 (1.0 - 3.3)†
Urban residence	1.0 (0.7 - 1.5)	1.3 (0.7 - 2.4)
Distance to nearest health facility		
<1 km	REF	REF
1-3 km	1.0 (0.6 - 1.4)	1.3 (0.8 - 2.2)
>3km	1.4 (0.7 - 2.7)	1.5 (0.6 - 3.7)

† $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; OR = odds ratio; CI = confidence interval; REF = reference group; km = kilometer

Note: Presence of danger signs excluded from final model due to collinearity with other covariates. Confidence intervals estimated using robust standard errors.

Supplementary Table 4.3. Multiply imputed results for interaction between illness severity and maternal employment on care-seeking for childhood illness between July 2015 and February 2016, rural Pune district, India

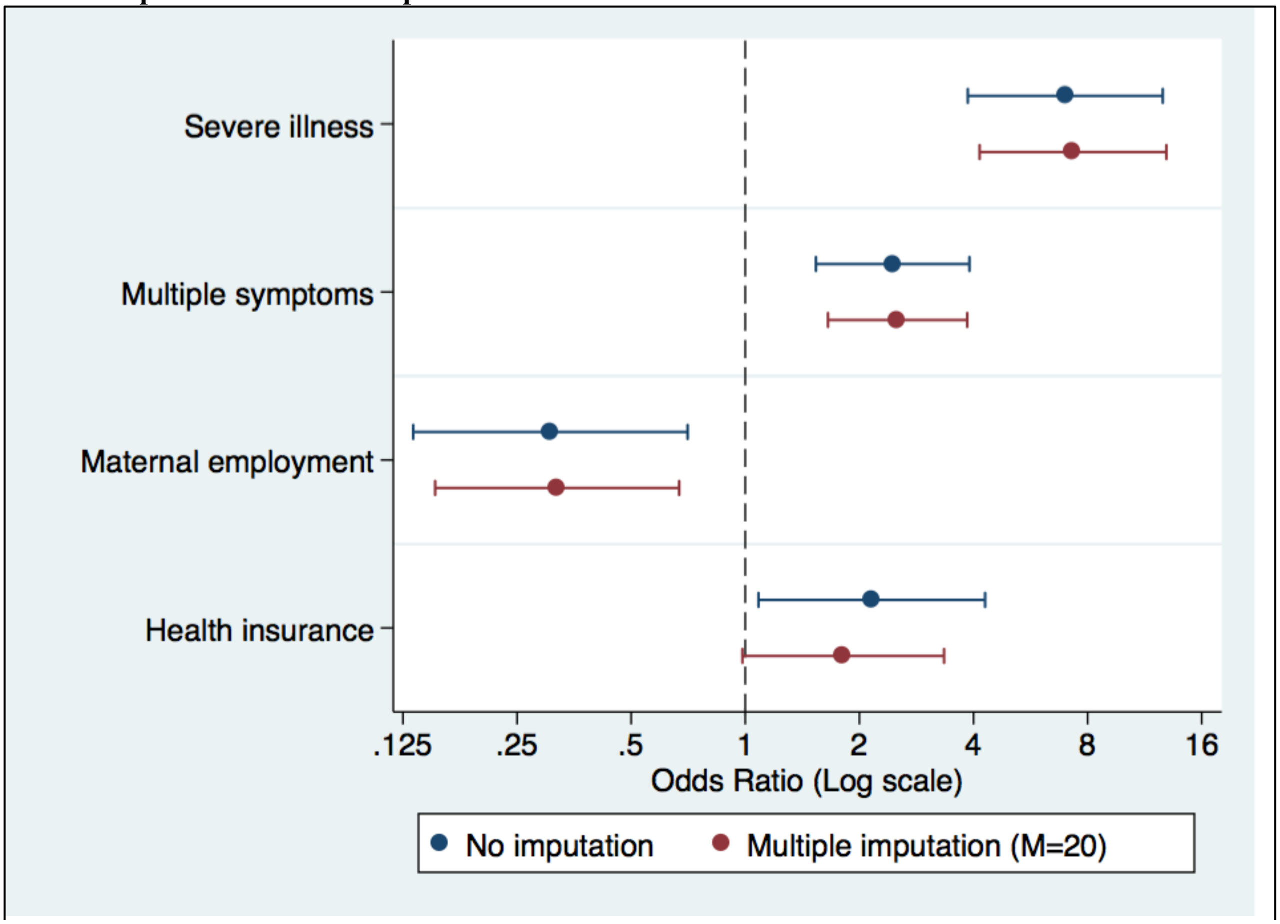
	Illness severity		
	Non-severe OR (95% CI)	Moderate-to-very severe OR (95% CI)	*Effect of severity within employment strata* OR (95% CI)
Maternal employment			
Not currently employed	1.0 (Reference)	7.3 (4.2 - 12.9)**	7.3 (4.2 - 12.9)**
Currently employed	0.3 (0.2 - 0.7)**	4.6 (2.1 - 9.8)**	14.3 (5.9 - 34.6)**
Effect of employment status within severity strata	0.3 (0.2 - 0.7)**	0.6 (0.3 - 1.4)	

† $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; OR = odds ratio; CI = confidence interval

Measure of interaction on multiplicative scale: ratio of ORs (95% CI) = 2.0 (0.7 - 5.6)

Note: ORs are adjusted for number of reported symptoms, maternal education, child age, child sex, household structure, religion of head of household, household SES, health insurance coverage, urban residence, and distance to nearest health facility. Confidence intervals estimated using robust standard errors.

Supplementary Figure 4.1: Comparison of selected regression results from multivariable model without imputation and with imputation



Chapter 5: TrackCare: A Smartphone-based Approach for Tracking Participant Movement in Rural Pune District, India

(Under review by the Journal of Global Health)

Andrew Marsh MPH[1, 2], Tathagata Bhattacharjee MBA[2, 3], Siddhivinayak Hirve PhD[2], Pallavi Lele MSc[2], Uddhavi Chavan MSc[2], Sandeep Bhujbal MCS[2, 3], Howard Lin PhD[4], Harish Nair PhD[5] Sanjay Juvekar PhD[2, 3†], Harry Campbell MD[5†]

[†] Joint last-author with equal contribution

[1] Institute for International Programs, Johns Hopkins University Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD, USA

[2] KEM Hospital Research Centre, Sardar Moodliar Road, Rasta Peth, Pune 411011, India

[3] INDEPTH Network, 40 Mensah Wood Street, East Legon, Accra, Ghana

[4] School of Informatics, University of Edinburgh, 10 Crichton St, Edinburgh EH8 9AB, United Kingdom

[5] Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, United Kingdom

Abstract

Background:

Common approaches to measure health behavior depend on the validity of participant responses and may be biased due to poor recall or non-completion. Technology-based approaches, such as those using Global Positioning System (GPS) data, provide an alternative approach while reducing these biases. This paper describes the development and implementation of TrackCare, a location-aware smartphone application used to detect maternal care-seeking for childhood illness.

Methods:

Mothers having at least one child < 5 years were enrolled and provided a GPS-enabled smartphone preinstalled with TrackCare, which recorded device location each minute and transferred data hourly to a central server. Household sociodemographic information was collected at baseline and field workers visited mothers monthly over six months to assess compliance with phone-related procedures. Location data completeness and quality were evaluated prospectively. Separate analyses evaluated the associations between various explanatory variables, data completeness, and participant compliance with study procedures, respectively.

Findings:

We enrolled 200 mothers and completed 1,092 follow-up interviews. Participant smartphones submitted location data for a mean of 152 days, representing 84% data completeness across the study duration. Compliance with study phone-related study procedures was 79% overall and significantly associated with increased data completeness ($\boldsymbol{\beta}_{\boldsymbol{Visits}\ \mathbf{1-3}}$ = 8%; 95% confidence

interval [CI] 3-13%; $\boldsymbol{\beta}_{Visits\ 4-6}$ = 19%; 95% CI 10-28%). Urban residence was associated with decreased data completeness ($\boldsymbol{\beta}$ = -4%; 95% CI -9% – -0.3%). Compliance was highest among participants of higher socioeconomic status (adjusted odds ratio [AOR] = 1.67; 95% CI 1.13 – 2.46) and increased during the second half of the study (AOR = 1.66; 95% CI 1.23 – 2.24). No association was observed between compliance and maternal age, education, baseline perceptions, or previous smartphone ownership.

Conclusions:

GPS-enabled smartphones provide a promising approach for the continuous, real-time measurement of movement data with increased data completeness relative to traditional GPS devices**.** These data provide useful insights into a range of participant health behaviors, such as care-seeking for childhood illness, with the potential to enhance data collection methods as approaches to analyze these location data are refined.

Introduction

An estimated 5.9 million children under the age of five die each year globally, with many of these deaths preventable through effective, low-cost interventions [1,2]. India has experienced significant gains in providing curative care for childhood illness with three quarters of children with suspected pneumonia and half of children with diarrhea receiving treatment [61]. However, further improvements are required to reach universal coverage. Advancing the coverage of these interventions within national programs requires valid approaches to monitor their delivery and use in the population [103].

Common approaches to measurement of intervention coverage include surveys and direct observation, all of which have their limitations. However, each of these approaches is subject to biases that decrease their validity and increase the uncertainty of estimates of their impact on mortality and morbidity. Surveys are dependent on the validity of participant responses and may be subject to biases due to poor recall or non-completion, while direct observation of health behavior is both costly and raises issues of privacy [27].

Technology-based approaches, such as those based on Global Positioning System (GPS) data, have the potential to provide a minimally invasive approach to monitor participant behavior continuously without the biases associated with traditional observational methods [29]. GPS receivers communicate with orbiting satellites to determine their location with a precision of up to a few meters, depending on the number of visible satellites and their relative placement in the sky [30]. GPS has traditionally been measured using dedicated devices, though it is suggested that the single function of these devices makes them less attractive to study participants and results in decreased compliance with study protocols [31].

Recently GPS receivers have become commonplace in many commercially available smartphones and have been shown to produce comparable data to traditional GPS devices across various settings [32-37]. GPS-enabled smartphones have not only been applied extensively in the field of transportation to model travel mode [38] and route choice [39], but also used to characterize participant life space [40] and social engagement [41] and to describe the context of participant activities, such as environmental exposure to air pollution [42] and hospitalizations. [43] Traditional GPS signal acquisition methods operate independently of the cellular network, enabling location estimation even when no cellular signal is available. Where coverage exists, smartphones use cellular network data to supplement traditional methods and determine a location fix more quickly than GPS receivers [34].

Several studies identified issues related to variable quality among different smart phone models [33,34,37], highlighting the importance of device selection and testing prior to study implementation. Additionally, smartphones permit real-time data transfer and communication with research participants, through device customizability, including the installation of additional applications and the configuration of device settings, introduces the potential for non-standard device performance [28]. It is possible to restrict these functions, but doing so may reduce attractiveness of the device to participants [45].

While GPS devices present a promising approach to study mobility and human behavior, there are limitations to their use. Krenn and colleagues observed frequent data loss in a review of 24 studies applying GPS to the study of physical activity with longer study duration associated with increased data loss [46]. Additionally, GPS performance is influenced by features of both the natural and built environment and may deteriorate in areas where large buildings obstruct satellite visibility [29]. Previous applications of GPS-enabled smartphones have predominantly occurred in developed countries and little is known about whether a similar approach would be feasible in a less-developed context.

As part of a care-seeking study for childhood illnesses in rural Pune district, India, we developed a location-aware smartphone application (TrackCare) to track mothers' movement and detect visits to local health facilities [85]. Data collected through this approach were compared to maternal report of care-seeking for childhood illness. This paper describes the development and implementation of the TrackCare application. Specific objectives included: 1) evaluating the completeness and quality of location data collected by the application, 2) assessing participant compliance with phone-specific study procedures, and 3) exploring the individual, household, and environmental factors associated with data completeness and participant compliance.

Methods

Pilot Testing and Device Selection

The study took place in the 22 villages of the Vadu Health and Demographic Surveillance System (HDSS), located 30 km northeast of the Pune city, Maharashtra state, India. Final study procedures were defined after pilot testing three phone models with a convenience sample of 29 participants from urban and rural villages. Pilot models were selected based on technical specifications (e.g. Android operating system, GPS chipset, battery capacity, dual SIM) and logistical concerns (e.g. availability for bulk purchase and local capacity to provide technical support). Selected models included Sony Xperia E1, Samsung Galaxy Core 2, and Micromax Unite 2 A106. Models were evaluated according to the location data completeness and quality and device battery life. Location data completeness and quality were highest among the Sony and Samsung models, with the Sony model having marginally higher battery life. Local device availability changed during piloting as manufacturers replaced previously selected models with newer versions, resulting in three additional models being evaluated in structured field-worker testing (Sony Xperia E3, Sony Xperia E4, and Samsung Galaxy S Duo 3). Final testing led to the selection of the Sony Xperia E4 model for use in the study.

Participant Recruitment and Follow-up

Participants were a random sample of mothers with young children drawn from a population database maintained by the Vadu HDSS. Inclusion criteria required mothers to be aged 15-49 years, have at least one child less than 60 months old at enrollment, and self-identify as the primary caregiver. After enrollment, participants were randomly assigned to one of three study groups: phone group, longitudinal comparison group, or cross-sectional comparison group. Given this study's objectives, this paper includes only those participants in the phone group. Study implementation was led by a team of five field workers with primary support and oversight from two researchers from the Vadu HDSS main office.

After randomization, participants assigned to the phone group were provided with a Sony Xperia E4 Dual SIM phone, a protective case, and a one-page information leaflet summarizing key features and usage instructions. Participants were asked to charge the phone daily, keep it on throughout the day, and carry the phone whenever seeking care. Participants were also provided with a SIM card enabled for unlimited mobile data usage and in-network communication. Field workers recorded the geolocation of each participant's household at baseline using Garmin e-Trex devices and administered a baseline questionnaire collecting individual and household-level sociodemographic information. Additional questions measured participant attitudes toward phone-specific elements of the study protocol.

Field workers conducted two support visits at three days and 11-14 days after enrollment, respectively, to identify and resolve any phone-related issues. Participants were then visited monthly over six months and administered a questionnaire primarily collecting information about care-seeking for childhood illness and including a supplementary module on current phone settings and four participant-reported phone behaviors, two of which were desirable (phone use, phone charging) and two of which were not (lending phone to others, experiencing problems with phone). Field workers identified and corrected any deviations from optimal settings and retrained mothers on correct phone use as necessary.

GPS Data Collection, Cleaning, and Processing

Participant GPS data were obtained by TrackCare, an Android application developed for this study and preinstalled on participant phones (Figure 5.1). TrackCare was designed using Android Studio to run on Android OS version 5.0 [104]. The application continually received location updates from the operating system, the frequency of which depended on GPS and cellular network signal strength. Each minute the phone recorded its current location (latitude, longitude, and positional accuracy), source of location data (e.g. GPS, mobile network provider), and the current configuration of the device's location settings. Records also included a timestamp indicating when the coordinate was saved. These records were stored internally on the phone device in an encrypted database and transmitted hourly to a central study server. When hourly transfer was not possible (e.g. due to poor network availability), cumulated data were transferred during the next scheduled transfer. Internally stored data were deleted from the device once successfully transmitted to the central server.

TrackCare included additional features to promote its continued functioning during the six-month observation period. The application required no participant input to run and included a minimal interface for field worker debugging. Additionally, the application would start automatically when the device was turned on and would restart itself if ever closed. Lastly, TrackCare password-protected its uninstallation, preventing removal except by study staff.

Participant data were stored on a server at the University of Edinburgh and mirrored on a server maintained at Vadu HDSS. An automated script checked these data daily for completeness and quality, generating a list of participants with a high proportion of missing location data, data indicating that location settings had been changed, or data indicating extended travel outside the study area. These participants were flagged for field worker follow up either by phone call or in-person visit.

The complete dataset of participant GPS data was cleaned and processed at study completion. Data cleaning consisted of removing duplicate records and adjusting incorrect coordinate timestamps. Duplicated coordinates may represent periods of no movement or periods where the device was unable to obtain a new location estimate and recycled a previously cached coordinate. We dropped all such duplicated coordinates with the understanding that some these coordinates represent valid periods of no movement. Data processing dropped outliers based on impossible speed and records with uncertainty around the measured coordinate exceeding 50 meters (i.e. positional accuracy > 50m) [105]. When multiple coordinates were recorded during the same minute, the best coordinate was selected using supplementary information. (e.g. source and location mode) Missing location data were imputed when neighboring coordinates were less than one hour apart or within 100 meters [106]. When imputed location data replaced coordinates that had previously been dropped (e.g. repeated coordinates), we calculated the distance between the original and imputed coordinates.

Analysis

This paper presents the results of two analyses. First, we evaluated the association between data completeness, participant compliance, and environmental factors. Second, we evaluated the association between participant compliance and individual and household-level participant covariates. The outcome variables in these analyses were data completeness and participant compliance, respectively. Data completeness was calculated as the proportion of time during the 14 days before each follow-up visit for which participant location data were available. A participant was classified as compliant at a follow-up visit if all the following conditions were met: phone available at visit and turned on, TrackCare app installed, and optimal settings configured (i.e. mobile data turned on, battery-saving mode turned off, and location settings set to “high accuracy”). Not meeting one or more criterion resulted in a classification of noncompliant at the follow-up visit.

Explanatory variables were selected based on their previous associations in the literature with additional variables included after conducting exploratory data analysis. We examined a different set of explanatory variables in each of the two analyses. The analysis of data completeness included compliance, reported behavior, residence, distance to highway, and study phase (i.e. visit 1-3 or 4-6). The analysis of compliance included maternal age, education, current employment status, baseline attitude toward phone-related procedures, socioeconomic status (defined as quintiles following the approach in [86]), religion, household structure (i.e. nuclear or joint), previous smartphone ownership, intention to use the study phone as the primary phone, residence, administrative block, distance to highway, and study phase. Distance to the highway was calculated in ArcGIS 10.2 [88].

Multiple linear regression was used to estimate the association between data completeness, participant compliance, and environmental factors, while multiple logistic regression was used to estimate the association between participant compliance and participant and household-level characteristics. Final model selection was guided by the results of bivariate analyses with Wald tests used to compare extended and null models. Standard errors in both regressions were adjusted to account for clustering at the participant level. All analyses were conducted in Stata, version 14 [90].

Ethical Considerations

Written consent was provided by all participants prior to randomization. Prior to obtaining consent, participants were informed that those assigned to the phone group would be given a smartphone and would be allowed to keep the device after involvement in the study regardless of completion. Confidentiality of participant location data was ensured throughout the study by encrypting all location data stored locally on participant devices and by erasing these data once transferred to the study server. The study protocol was approved by the ethics committees of the University of Edinburgh and K.E.M. Hospital Research Centre, Pune (Study ID No. 1415).

Results

Field workers enrolled 200 participants from June to September 2015, baseline characteristics for whom are presented in Table 5.1. One enrolled participant withdrew before the first follow-up visit and was excluded from analysis. Average participant age was 25.3 years (SD 3.3) with 10.9 completed years of education (SD 2.7). Most participants were housewives (69%), followed by employment in agriculture (19%). Two thirds of households were located in urban villages with a third of households located 250 meters or less from the main highway. The majority of participants resided in Taluka Shirur, the more populous of the two administrative zones included in the Vadu HDSS study area.

Participant reported perceptions regarding phone-related study procedures varied by item. Less than a third of participants reported any concern regarding another family member taking possession of the phone, carrying the phone, potential theft of the phone, or potential damage to the phone. About half of participants reported some concern about their safety when carrying the phone or about potentially losing the phone. Many participants expressed concern regarding their ability to charge the phone each day, keep it on at all times, and carry it with them when seeking care for a sick child. Two thirds of households reported previous smartphone ownership with all but one household reporting an intention to use the study phone as the household's primary mobile phone.

Participant mobile phones submitted coordinates corresponding to a mean observation time of 151.8 days (SD 24.7), which constituted 84% of the six-month follow-up period. Figure 4.2 illustrates the steps by which these participant data were cleaned and processed. After removing duplicate coordinates and unreliable data, median accuracy was 12 meters (IQR 9-22) with 53% of all coordinates transferred to the study server within one hour of being recorded. Supplementary Tables 5.1-5.3 include additional details on the quality of the data used for imputation, the timeliness with which data were transferred to the study server, and the distance between imputed points and those they replaced.

Table 5.2 presents summary data on follow-up visits by completion status. Participants completed 1,092 of 1,194 planned follow-up visits (9% loss to follow-up). Compliance data are available for 977 (89%) of these completed follow-up visits, which constitutes the evaluable sample in the remaining analyses.

Table 5.3 presents participant data completeness, compliance, and reported phone behaviors during the study. Median GPS data completeness across all visits was 91% (IQR 78-97%) with minimal variability between follow-up visits. Participants met all compliance criteria at 79% of follow-up visits, with 15% of participants classified as noncompliant because optimal phone settings had been changed and 5% of participants classified as compliant because the study phone was unavailable during the follow-up visit and phone settings could not be observed. The most frequent changes to phone settings were disabling mobile data (7%), turning off the phone (3%), and disabling "High Accuracy" mode within the location settings (3%). Compliance status could not be determined for 2% of follow-up visits where the study phone was available but data on one or more phone settings were missing. Participants reported high frequency of each desirable behavior (>90%) and low frequency of each undesirable behavior (<10%) across all study visits.

Table 5.4 presents the results of simple and multiple linear regression on data completeness. The unadjusted analysis found that participant compliance, frequent reported phone use, and frequent reported phone charging were each individually associated with an increase in data completeness, while urban residence and later study visits were associated with decreased data completeness. The final model of data completeness includes participant compliance, urban residence, study phase, and an interaction term between compliance and study phase (Table 5.5). While reported phone use and phone charging were also associated with increased data completeness, these were omitted from the final model due to high correlation with participant compliance. Compliance was associated with an 8% increase in data completeness during visits 1-3 and a 19% increase in data completeness during visits 4-6, both of which were highly statically significant ($p < 0.01$). Conversely, data completeness decreased by 13% from visits 4-6 relative to visits 1-3 among those who were not compliant with study procedures ($p < 0.01$) while no significant difference was observed between visits 4-6 and visits 1-3 among those who were compliant. Urban residence was associated with a 4% decrease in data completeness.

Table 5.6 presents the results of the unadjusted and adjusted analyses of participant compliance. Significant unadjusted associations were observed between compliance and concern about damaging the phone, household SES, distance from the highway, and study phase. Marginally significant associations were observed between concern about keeping the phone turned on and religion. The final model included household SES, administrative block, distance from the highway, and study phase. Maternal concern about phone damage was considered but resulted in no significant increase in model fit. Compliance was highest among households in the highest SES quintiles (OR 1.67; 95% CI 1.13-2.46) and during the final three follow-up visits. (OR 1.66, 95% CI 1.23-2.24) Compliance was marginally higher among households at least 250 meters from the highway. (OR = 1.41, 95% CI 0.96-2.09)

Discussion

Our study deployed a location-aware application on GPS-enabled smartphones to collect fine-scale movement data in rural Pune district, India, with an average observation period of 151.8 days across 199 participants. To our knowledge, this is the largest study to collect movement data by smartphone and the first such study in a developing country to do so over an extended time period.

We compare these results with two notable examples. The Human Mobility Project provided its application for free online download, enrolling 270 participants from 13 countries and collecting an average of 6.3 days of location data each [44]. Glasgow and colleagues provided a smartphone application to 42 participants, collecting an average of 122.7 days of observation each [42]. A shared feature of these two studies is their use of participants' currently owned smartphones for their applications, greatly reducing study implementation cost. We considered deployment on participant-owned phones but concerns about inter-device variability led us to select a single device. High cost of supplying participants with mobile devices and mobile data plans may partially explain why fewer studies have attempted implementing this approach at large scale. Within our study context, mobile data plan costs were relatively minor in comparison with device costs and facilitated an extended follow-up period with minimal incremental phone-related costs.

Previous studies have identified data completeness as a key concern among GPS-based studies with missing data up to 92% [46]. In contrast, our study experienced relatively high data completeness of 84%. Increased data completeness is consistent with other studies collecting these data by smartphone [42], providing further support for this approach relative to traditional GPS devices. Data quality, measured as positional accuracy, is similar to studies evaluating smartphone performance in controlled conditions [32-34], though significantly better than what was observed in studies where participants' own devices are used [42,44]. We suspect this is due to a combination of inter-device variability and device configuration, especially the currently enabled location mode. Participants in our study were asked to maintain location mode in "High Accuracy" mode, which uses both available sources (GPS receiver and mobile network) to determine the best location estimate. We collected each phone's currently configured location mode with each record and flagged for follow up any participants changing their setting. Without this guidance it is foreseeable that participants wishing to increase battery life would enable "Battery Saving" mode. This mode obtains location estimates exclusively through the less energy-intensive mobile network, disabling the more accurate GPS receiver.

We explored factors contributing to data completeness and participant compliance with a subset of phone-related study procedures. Completeness was highest among participants complying with study procedures with no significant change between study visits. In contrast, non-compliance was associated with an 8% decrease in data completeness during the first three study visits and a 19% decrease during the final three study visits. This increased effect size is potentially concerning for studies conducted over extended periods and merits further investigation. Decreased data completeness among urban participants may be due to generally poorer performance of GPS in urban environments, where satellite signal is frequently obstructed or distorted by large buildings [30].

Compliance was high overall but increased significantly during the second half of the study, perhaps resulting from continued field worker support throughout the extended study period. We also noticed increased compliance among participants of higher SES. While increased compliance among participants of higher SES could feasibly result from increased exposure to smartphones, we saw no association between previous smartphone ownership and compliance. Alternatively, higher SES may indicate greater overall exposure to technology and consequently result in increased compliance.

No association was detected between compliance and maternal age, education, baseline perceptions, or previous smartphone ownership. This should be interpreted alongside the sample size, which was calculated to address the parent study objective. Our findings rule out a large effect size due to these variables but may still be consistent with the presence of a smaller effect size. This suggests that a technology-based approach with regular field worker support may be accessible to a broad range of participants with regard to education level, previous experience with technology, or baseline perceptions.

We faced several challenges in our approach. First, the large volume of data generated each day limited our ability to respond to all identified issues in a timely fashion. This was highlighted during one instance of widespread network unavailability, which resulted in delayed data transfer across all participants. Data were transferred when network connectivity returned but were flagged as potentially missing during the interim. Additionally, one participant lived beyond mobile network range and required frequent field worker visits to facilitate transfer of location data. Second, we experienced a few instances when the TrackCare application stopped transmitting coordinates and we were unable to retrieve locally stored location data. Such instances were resolved by reinstalling the application but resulted in some periods of data loss. Third, location data collected by the phone may misclassify participants' true location during periods when the phone was not carried or when the phone was carried by someone else. This potential misclassification is common to GPS-based studies, though the overall impact may be small. Isaacson and colleagues conducted an initial study into participant compliance with carrying the GPS device and found that participants forgot to carry the device during less than 5% of total observation time [107]. Fourth, compliance is measured according to a subset of phone-related behaviors that were readily observable at scheduled follow-up visits, though this measurement omits key components of compliance (e.g. charging behavior, lending the phone to others) and may not be representative of other time periods. Finally, while our study benefits from a large sample relative to other smartphone-based GPS studies, sample size may be insufficient to detect the effect of several individual and household characteristics.

The results of this study provide further evidence for the potential use of a smartphone-based approach for real-time participant movement tracking. Furthermore, our study suggests the suitability of this approach in a developing country with strong mobile phone infrastructure and high ownership base. Household mobile phone ownership in India exceeds 90% both nationally and within the state of Maharashtra [61]. As mobile phone ownership increases globally, the range of settings where such an approach may be implemented will continue to expand. These data provide useful insight into a range of participant health behaviors, such as care-seeking for childhood illness. As these applications are refined to accurately detect health facility visits in real-time, they may be extended to include a short survey after a visit occurs, collecting data on service provision and quality of care with minimal recall bias.

Conclusion

GPS-enabled smartphones provide a promising approach for the measurement of fine-scale movement data over extended periods. Furthermore, real-time data transmission allows for continuous monitoring of data quality and resolution of identified issues. GPS-enabled smartphones, coupled with routine field worker support, provide increased data completeness relative to studies using traditional GPS devices. Data completeness also depends upon participant compliance with several phone-related procedures, such as keeping the phone charged and turned on. Researchers interested in implementing such an approach should give special consideration to what mechanisms may be available to monitor and promote participant compliance throughout the study duration.

Funding

The study was funded by a prime award by the Bill and Melinda Gates Foundation (grant no. OPP1084442) made to the Johns Hopkins University of which University of Edinburgh, Edinburgh, UK and KEM Hospital Research Centre, Pune, India were sub-grantees.

Acknowledgements

Vadu HDSS is a member of the INDEPTH Network. We are grateful to the members of the Improving Coverage Measurement Core Group for their guidance in developing this study. We also thank Drs. Robert Black, Joanne Katz, Krishna Rao and Melinda Munos of the Johns Hopkins School of Public Health for their comments on the study protocol and this paper. Lastly, we wish to thank the study participants and the project staff at the Vadu HDSS, without whom this study would not have been possible.

Author Contributions

All authors conceived of the study and protocol. HL developed the TrackCare source code. AM, TB, PL, UC, SB, HL, and SJ conducted data collection. AM analyzed the data and wrote the paper. TB, SH, HN, SJ, and HC provided guidance on the analysis and interpretation of results. All authors read and agree with the manuscript and conclusions.

Competing Interests

The authors declare no competing interests.

References

1 Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet (London, England). 2016;388:3027-35.

2 Jones G, Steketee RW, Black RE, Bhutta ZA, Morris SS. How many child deaths can we prevent this year? The Lancet. 2003;362:65-71.

27 Kroeger A. Response errors and other problems of health interview surveys in developing countries. World health statistics quarterly Rapport trimestriel de statistiques sanitaires mondiales. 1985;38:15-37.

28 Maas J, Sterkenburg RP, de Vries SI, Pierik F. Using GPS to Measure the Interaction Between Individuals and Their Neighbourhood. Neighbourhood Structure and Health Promotion2013. p. 153-75.

29 Kerr J, Duncan S, Schipperijn J. Using global positioning systems in health research: a practical approach to data collection and processing. Am J Prev Med. 2011;41:532-40.

30 Enge P, Misra P. Special issue on Global Positioning System. Proceedings of the Ieee. 1999;87:3-15.

31 Paz-Soldan VA, Reiner RC, Jr., Morrison AC, Stoddard ST, Kitron U, Scott TW, et al. Strengths and weaknesses of Global Positioning System (GPS) data-loggers and semi-structured interviews for capturing fine-scale human mobility: findings from Iquitos, Peru. PLoS Negl Trop Dis. 2014;8:e2888.

32 Michael Y, McGregor EM, Allen J, Fickas S. Observing Outdoor Activity Using Global Positioning System-Enabled Cell Phones. In: Helal S, Mitra S, Wong J, Chang CK, Mokhtari M, editors. Smart Homes and Health Telematics: 6th International Conference, ICOST 2008 Ames, IA, USA, June 28-July 2, 2008 Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 177-84.

33 Wu J, Jiang C, Liu Z, Houston D, Jaimes G, McConnell R. Performances of different global positioning system devices for time-location tracking in air pollution epidemiological studies. Environ Health Insights. 2010;4:93-108.

34 Zandbergen PA, Barbeau SJ. Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones. Journal of Navigation. 2011;64:381-99.

35 Benson AC, Bruce L, Gordon BA. Reliability and validity of a GPS-enabled iPhone "app" to measure physical activity. Journal of sports sciences. 2015;33:1421-8.

36 Garnett R, Stewart R. Comparison of GPS units and mobile Apple GPS capabilities in an urban landscape. Cartography and Geographic Information Science. 2015;42:1-8.

37 Jahkola O, Happonen A, Knutas A, Ikonen J, editors. What should application developers understand about mobile phone position data. ACM International Conference Proceeding Series; 2017.

38 Wu LL, Yang B, Jing P. Travel Mode Detection Based on GPS Raw Data Collected by Smartphones: A Systematic Review of the Existing Methodologies. Information. 2016;7:67.

39 Kazagli E, Chen J, Bierlaire M. Individual mobility analysis using smartphone data. Mobile Technologies for Activity-Travel Data Collection and Analysis2014. p. 187-208.

40 Wan N, Lin G. Life-space characterization from cellular telephone collected GPS data. Computers Environment and Urban Systems. 2013;39:63-70.

41 Brusilovskiy E, Klein LA, Salzer MS. Using global positioning systems to study health-related mobility and participation. Social Science & Medicine. 2016;161:134-42.

42 Glasgow ML, Rudra CB, Yoo EH, Demirbas M, Merriman J, Nayak P, et al. Using smartphones to collect time-activity data for long-term personal-level air pollution exposure assessment. J Expo Sci Environ Epidemiol. 2016;26:356-64.

43 Nguyen KT, Olgin JE, Pletcher MJ, Ng M, Kaye L, Moturu S, et al. Smartphone-Based Geofencing to Ascertain Hospitalizations. Circulation Cardiovascular quality and outcomes. 2017;10.

44 Palmer JR, Espenshade TJ, Bartumeus F, Chung CY, Ozgencil NE, Li K. New approaches to human mobility: using mobile phones for demographic research. Demography. 2013;50:1105-28.

45 Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res. 2013;15:e86.

46 Krenn PJ, Titze S, Oja P, Jones A, Ogilvie D. Use of global positioning systems to study physical activity and the environment: a systematic review. Am J Prev Med. 2011;41:508-15.

61 International Institute for Population Sciences (IIPS) and ICF. National Family Health Survey (NFHS-4), 2015-16: India. Mumbai: IIPS, 2017.

85 Hirve S, Marsh A, Lele P, Chavan U, Bhattacharjee T, Nair H, et al. Concordance between GPS-based smartphone app for continuous location tracking and mother's recall of care-seeking for child illness in India. 2018.

86 Filmer D, Pritchett LH. Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India. Demography. 2001;38:115-32.

88 ESRI. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute; 2011.

90 StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP; 2015.

103 Requejo JH, Newby H, Bryce J. Measuring coverage in MNCH: challenges and opportunities in the selection of coverage indicators for global monitoring. PLoS Med. 2013;10:e1001416.

104 Android Developers. Android Studio. Available: https://developer.android.com/studio/. Accessed: May 15, 2018.

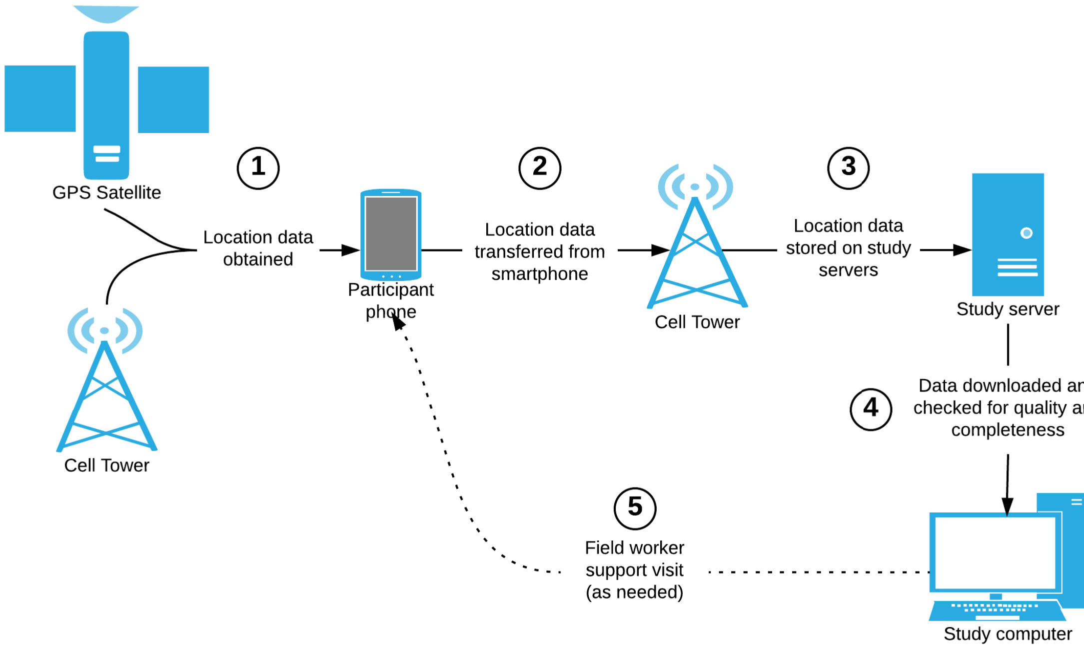
105 Vich G, Marquet O, Miralles-Guasch C. Suburban commuting and activity spaces: using smartphone tracking data to understand the spatial extent of travel behaviour. Geographical Journal. 2017;183:426-39.

106 Shareck M, Kestens Y, Gauvin L. Examining the spatial congruence between data obtained with a novel activity location questionnaire, continuous GPS tracking, and prompted recall surveys. Int J Health Geogr. 2013;12:40.

107 Isaacson M, Shoval N, Wahl HW, Oswald F, Auslander G. Compliance and data quality in GPS-based studies. Transportation. 2016;43:25-36.

Tables and Figures

Figure 5.1 TrackCare Process Flow



Note: (1) TrackCare continuously communicated with available satellites and cell towers to obtain location data, saving these data locally on participant smartphones. (2) These data were transferred hourly over the mobile data network and (3) stored on servers maintained by the University of Edinburgh and Vadu HDSS. (4) Study staff downloaded these data each day to check their completeness and quality. (5) Participants not meeting thresholds for data completeness or quality were visited by field workers to identify to support phone-related study procedures.

Table 5.1 Baseline participant and household characteristics (n = 199)

Characteristic	n (%)
Participant variables	
Maternal age, mean (SD)	25.3 (3.3)
Maternal education, completed years (mean 10.9, SD 2.7)	
5-7 years	22 (11.5%)
8-9 years	30 (15.7%)
10-11 years	55 (28.8%)
12+ years	84 (44.0%)
Maternal employment status	
Agricultural employment	37 (19.3%)
Other employment	23 (12.0%)
Unemployed	132 (68.8%)
Perceptions towards phone (Agree/Strongly Agree)	
I worry about charging the phone every day.	129 (66.2%)
I worry about keeping the phone on during the day.	142 (72.8%)
I worry about carrying the phone with me when I seek treatment or advice for my sick child.	147 (75.4%)
I worry that someone may try to steal the phone.	62 (32.0%)
I worry that the phone will not be easy to carry.	49 (25.4%)
I worry that I will lose the phone.	99 (50.8%)
I worry that I or someone in my household will damage the phone.	63 (32.5%)
I worry about my safety when I carry the phone.	96 (49.2%)
I worry that another family member will take the phone for himself or herself.	30 (15.4%)
Overall, I feel the phone will be easy to use.	164 (84.1%)
Household variables, non-environmental	
Religion of household head	
Hindu	183 (92.4%)
Other	15 (7.6%)
Household structure	
Nuclear	89 (45.9%)
Joint	105 (54.1%)
Intent to use as primary phone	
No	1 (0.5%)
Yes	194 (99.5%)
Previous smartphone ownership	
No	60 (31.2%)
Yes	132 (68.8%)
Household variables, environmental	
Residence	
Rural	62 (31.2%)
Urban	137 (68.8%)
Administrative block	
Shirur	140 (70.4%)
Haveli	59 (29.6%)
Distance to main highway	
Less than 250 meters	71 (35.6%)
Greater than or equal to 250 meters	128 (64.3%)

Figure 5.2 Processing of submitted GPS records

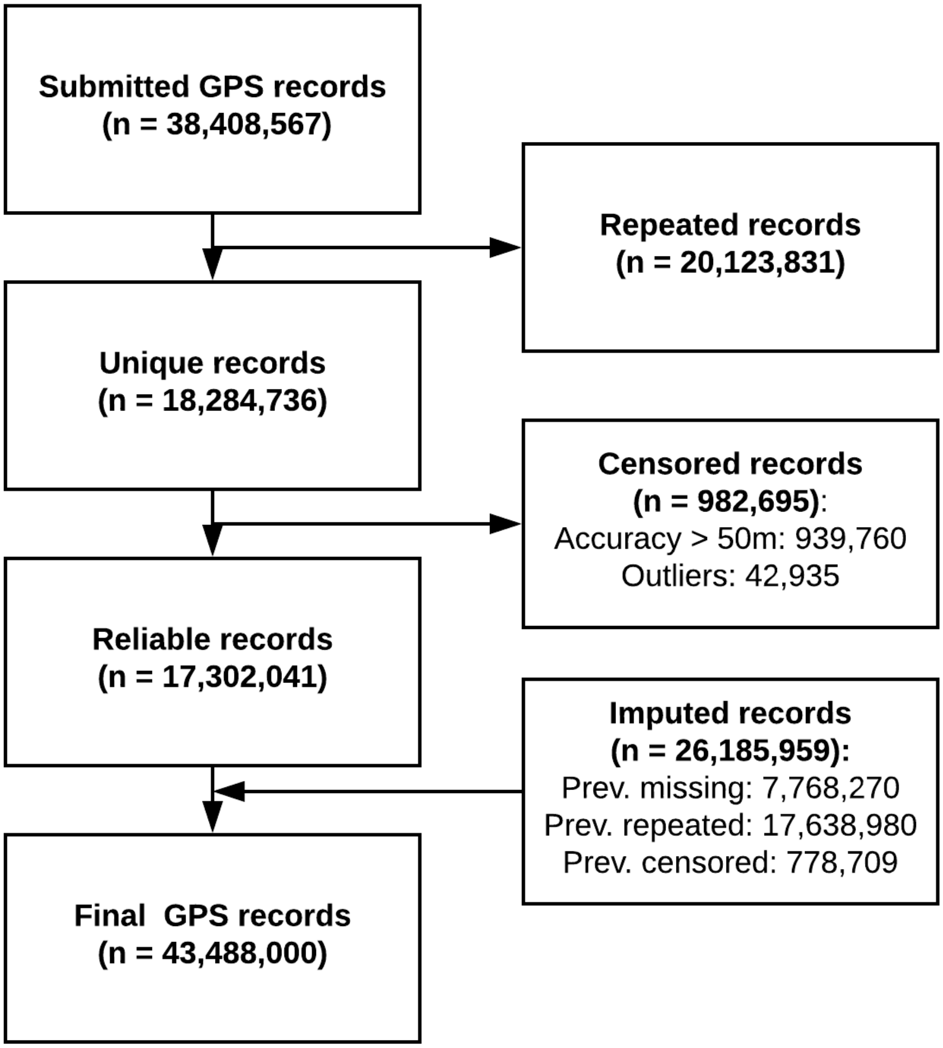


Table 5.2 Follow-up visits by completion status

	Visits planned	Visits Completed, n (% of planned)	
		All	Visits with compliance data
All Visits	1194	1092 (91%)	977 (82%)
Visit 1	199	188 (94%)	161 (81%)
Visit 2	199	188 (94%)	171 (86%)
Visit 3	199	183 (92%)	161 (81%)
Visit 4	199	180 (90%)	168 (84%)
Visit 5	199	174 (87%)	158 (79%)
Visit 6	199	179 (90%)	158 (79%)

Table 5.3 Data completeness, observed compliance, and reported participant behaviors

Characteristic	All Visits (n=977)	Visit 1 (n=161)	Visit 2 (n=171)	Visit 3 (n=161)	Visit 4 (n=168)	Visit 5 (n=158)	Visit 6 (n=158)
Data completeness, median (IQR)	91% (78-97%)	93% (79-97%)	92% (81-98%)	92% (79-97%)	90% (77-96%)	89% (80-98%)	89% (69-97%)
Observed compliance							
Compliant	768 (79%)	123 (76%)	120 (70%)	129 (80%)	140 (83%)	133 (84%)	123 (78%)
Noncompliant	194 (20%)	36 (22%)	51 (30%)	31 (19%)	24 (14%)	21 (13%)	31 (20%)
Phone unavailable at visit	48 (5%)	9 (6%)	10 (6%)	9 (6%)	6 (4%)	5 (3%)	9 (6%)
Phone available at visit, configuration not optimal	146 (15%)	27 (17%)	41 (24%)	22 (14%)	18 (11%)	16 (10%)	22 (14%)
Mobile data turned off	71 (7%)	10 (6%)	19 (11%)	14 (9%)	10 (6%)	7 (4%)	11 (7%)
Phone turned off	32 (3%)	5 (3%)	8 (5%)	5 (3%)	5 (3%)	5 (3%)	4 (3%)
Location services not set to "High Accuracy"	26 (3%)	6 (4%)	7 (4%)	4 (2%)	1 (1%)	4 (3%)	4 (3%)
Battery-saving ("STAMINA") mode enabled	22 (2%)	6 (4%)	8 (5%)	4 (2%)	2 (1%)	1 (1%)	1 (1%)
TrackCare application uninstalled	9 (1%)	2 (1%)	3 (2%)	1 (1%)	1 (1%)	0 (0%)	2 (1%)
Compliance status missing (i.e. phone available, configuration unknown)	15 (2%)	2 (1%)	0 (0%)	1 (1%)	4 (2%)	4 (3%)	4 (3%)
Reported behaviors (“Often” or “Always”), n (%)							
Use the phone	913 (94%)	143 (90%)	163 (96%)	157 (98%)	157 (93%)	149 (94%)	144 (91%)
Charge the phone	944 (97%)	156 (97%)	167 (98%)	155 (97%)	167 (99%)	149 (94%)	150 (95%)
Someone other than participant carried phone	31 (3%)	3 (2%)	10 (6%)	5 (3%)	1 (1%)	6 (4%)	6 (4%)
Had problems with the phone not working	12 (1%)	0 (0%)	1 (1%)	2 (1%)	1 (1%)	6 (4%)	2 (1%)

Table 5.4 Unadjusted and adjusted associations between compliance, environmental characteristics, and data completeness

Characteristic	Unadjusted β (95% CI)	Adjusted β (95% CI)
Intercept	n/a	0.82
Observed compliance	0.12 (0.07 – 0.18)**	0.08 (0.03 – 0.13)**
Reported Participant Behavior ("Often/always" vs. "Never/rarely/sometimes")		
How often did you use the phone?	0.15 (0.05 – 0.24)**	-
How often did you charge your phone?	0.17 (0.03 – 0.31)*	-
How often did someone other than you carry the phone?	0.01 (-0.04 – 0.06)	-
How often did you have problems with the phone not working?	-0.12 (-0.39 – 0.15)	-
Environmental Characteristics		
Urban residence	-0.04 (-0.09 – 0.030)*	-0.04 (-0.09 – 0.003)*
Administrative block (Haveli vs. Shirur)	-0.01 (-0.06 – 0.03)	-
Distance to highway, km	0.01 (-0.001 – 0.02)†	-
Study phase		
Visits 4-6 vs. Visits 1-3	-0.03 (-0.06 – -0.005)*	-0.13 (-0.21 – -0.05)**
Interaction Terms		
Observed Compliance * Visit, binary	n/a	0.11 (0.03 – 0.20)*

† $p < 0.10$; * $p < 0.05$; ** $p < 0.01$

Table 5.5 Effect modification of the association between compliance and data completeness by follow-up visit period

Characteristic	Adjusted β (95% CI)
Intercept	0.82
Compliance	
Visits 1-3	0.08 (0.03 – 0.13)**
Visits 4-6	0.19 (0.10 – 0.28)**
Visits 4-6 vs. Visits 1-3	
Compliance = 0	-0.13 (-0.21 – -0.05)**
Compliance = 1	-0.02 (-0.05 – 0.01)
Urban residence	-0.04 (-0.09 – 0.003)*

* $p < 0.05$; ** $p < 0.01$

Table 5.6 Unadjusted and adjusted associations between maternal and household variables on compliance

Characteristic	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Maternal age	1.02 (0.95 - 1.09)	-
Maternal education		
5-7 years	REF	-
8-9 years	1.27 (0.56 - 2.88)	-
10-11 years	1.37 (0.73 - 2.55)	-
12+ years	1.23 (0.70 - 2.16)	-
Maternal employment		
Currently employed	REF	-
Employed, agriculture	0.99 (0.54 - 1.83)	-
Employed, non-agriculture	1.50 (0.89 - 2.55)	-
Maternal perception toward phone use (agree/strongly agree)		
I worry about charging the phone every day.	0.82 (0.53 - 1.27)	-
I worry about keeping the phone on during the day.	0.65 (0.41 - 1.04)†	-
I worry about carrying the phone with me when I seek treatment or advice for my sick child.	0.70 (0.43 - 1.15)	-
I worry that someone may try to steal the phone.	1.19 (0.79 - 1.80)	-
I worry that the phone will not be easy to carry.	1.24 (0.81 - 1.92)	-
I worry that I will lose the phone.	0.79 (0.53 - 1.17)	-
I worry that I or someone in my household will damage the phone.	0.63 (0.41 - 0.95)*	-
I worry about my safety when I carry the phone.	0.97 (0.66 - 1.44)	-
I worry that another family member will take the phone for himself or herself.	0.97 (0.56 - 1.66)	-
Overall, I feel the phone will be easy to use.	0.86 (0.46 - 1.59)	-
Household Socioeconomic Status		
Quintiles 1-2	REF	REF
Quintiles 3-5	1.72 (1.17 - 2.54)**	1.67 (1.13 - 2.46)*
Religion of head of household		
Hindu	REF	-
Other	0.60 (0.34 - 1.06)†	-
Household Structure		
Nuclear	REF	-
Joint	1.20 (0.81 - 1.77)	-
Previous smartphone ownership	1.01 (0.66 - 1.54)	-
Urban residence	0.97 (0.63 - 1.5)	-
Distance from highway		
Less than 250 meters	REF	REF
Greater than or equal to 250 meters	1.50 (1.02 - 2.21)*	1.41 (0.96 - 2.09) †
Study phase		
Visits 1-3	REF	REF
Visits 4-6	1.65 (1.23 - 2.23)**	1.66 (1.23 - 2.24)**

† $p < 0.10$; * $p < 0.05$; ** $p < 0.01$

Supplementary Tables and Figures

Supplementary Table 5.1 Positional accuracy by location mode and location source for all coordinates used in interpolation

Location mode	Source	Number of points	Percent of total	Accuracy (meters), Median (IQR)
All location modes	All Sources	17,302,041	100.0%	11.7 (8.8-21.9)
	GPS	16,634,185	96.1%	11.3 (8.8-21.1)
	Network	480,032	2.8%	36 (27-40.5)
	Fused	187,824	1.1%	14.4 (9.6-24.9)
High accuracy	All Sources	17,020,299	98.4%	11.7 (8.9-21.9)
	GPS	16,363,258	94.6%	11.4 (8.8-21.2)
	Network	469,245	2.7%	36 (27-40.5)
	Fused	187,796	1.1%	14.4 (9.6-24.9)
Device only	All Sources	270,504	1.6%	9.6 (8.3-13)
	GPS	270,478	1.6%	9.6 (8.3-13)
	Network	-	-	-
	Fused	26	0.0%	28.8 (28.8-28.8)
Battery saving	All Sources	10,768	0.1%	37.5 (36-40.5)
	GPS	2	0.0%	9.1 (9.1-9.2)
	Network	10,764	0.1%	37.5 (36-40.5)
	Fused	2	0.0%	37.7 (36.4-39)
Location settings disabled	All Sources	470	0.0%	8.7 (8.7-8.7)
	GPS	447	0.0%	8.7 (8.7-8.7)
	Network	23	0.0%	26 (25-38)
	Fused	-	-	-

Supplementary Table 5.2 Time from saving coordinate on device to transfer to server (hours), for all coordinates used in interpolation

Time to submission	Number of points	Percent of total
Less than 1 hour	9,150,512	52.9%
1-3 hours	3,464,974	20.0%
3-6 hours	1,820,482	10.5%
6-12 hours	1,260,909	7.3%
12-24 hours	718,837	4.2%
1-3 days	548,320	3.2%
3-7 days	259,691	1.5%
More than 7 days	78,316	0.5%
Total	17,302,041	100.0%

Supplementary Table 5.3 Distance between dropped and interpolated coordinates by observation type

Observation type	Number of points	Percent of total	Distance (meters), Median (IQR)
All types	18,417,689	100%	8 (0-57)
Repeated value	17,638,980	96%	7 (0-42)
Dropped point, accuracy > 50m	735,800	4%	558 (219-1407)
Dropped point, outlier	42,909	1%	1191 (509-3225)

Chapter 6: Development and Implementation of a GPS-based Approach to Detect Care-seeking for Childhood Illness

Andrew Marsh MPH[1, 2], Siddhivinayak Hirve PhD[2], Pallavi Lele MSc[2], Uddhavi Chavan MSc[2], Tathagata Bhattacharjee MBA[2, 3], Harish Nair PhD[4] Sanjay Juvekar PhD[2, 3†], Harry Campbell MD[4†]

[†] Joint last-author with equal contribution

[1] Institute for International Programs, Johns Hopkins University Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD, USA

[2] KEM Hospital Research Centre, Sardar Moodliar Road, Rasta Peth, Pune 411011, India

[3] INDEPTH Network, 40 Mensah Wood Street, East Legon, Accra, Ghana

[4] Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, United Kingdom

Abstract

Background:

Common approaches to measure health behaviors rely on the validity of participant responses and are subject to bias. Technology-based alternatives, particularly those using GPS, address these biases while opening new channels for research. This study describes the development and implementation of a GPS-based approach to detect health facility visits in rural Pune district, India, nested within a study evaluating maternally reported care-seeking behavior for childhood illness.

Methods:

Participants were mothers of children under-five within the study area of the Vadu Health and Demographic Surveillance System. Participants received GPS-enabled smartphones preinstalled with a location-aware application, TrackCare, that continually recorded participant GPS data and transferred these data to a central server. Data were analyzed to identify health facility visits according to a parameter-based approach, optimal values of which were calibrated through locally simulated health facility visits. Participants reviewed all detected visits through monthly prompted recall surveys, confirming those visits which were correctly identified. Detected visits were analyzed using logistic regression methods to explore factors associated with the detection of false positive visits.

Findings:

Field workers enrolled 199 participants and completed 1,098 follow-up visits over the six-month study period. Prompted recall surveys were completed for 694 follow-up visits with at least one GPS-based health facility visit detected. While the approach performed well during calibration

(AUC 0.95), overall performance was low when applied to participant data with 440 of 22,251 detected visits confirmed (2%). False positives increased as participants spent more time in high health facility density areas (odds ratio [OR] 2.29, 95% confidence interval [CI] 1.62-3.25). Visits detected at facilities other than hospitals and clinics were also more likely to be false positives (OR 2.78, 95% CI 1.65-4.67). False positives were more likely among visits detected nearby participant homes with the likelihood decreasing as distance increased (OR 0.89, 95% CI 0.82-0.97). Visit duration was not associated with confirmation status.

Conclusions:

Field-worker simulated health facility visits resulted in an optimal parameter combination that performed well within the simulation dataset but substantially overestimated health visits when applied to participant GPS data. This study provides useful insight into the challenges in detecting health facility visits where providers are numerous and their location is highly clustered within urban centers.

Introduction

An estimated 5.9 million children under five die each year globally with pneumonia, malaria, and diarrhea among the leading causes [1]. Proven preventative and curative interventions exist to reduce mortality and morbidity from these causes yet their impact is limited by poor access to healthcare [2]. Data on various household health behaviors, including care-seeking for childhood illness, are commonly collected through large-scale household surveys, such as the Demographic and Health Survey and the Multiple Indicator Cluster Survey [75,77]. These data typically rely on maternal self-report, the validity of which is subject to various biases [108]. Technology-based approaches to collect information on participant behavior, especially those using Global Positioning System (GPS) sensors, provide an alternative to survey-based approaches, minimizing the biases inherent in those traditional approaches while broadening the range of topics which may be explored [29].

GPS has been increasingly applied within the field of health research. Depending on satellite visibility and location, GPS receivers are capable of continually determining their location with a precision of several meters [30]. Among the many applications of GPS are the study of disease exposure and transmission [109-111], environmental exposure [33,112], social interaction and exclusion [41,113], mobility-related illness outcomes [114], and physical activity [46]. While GPS has traditionally been measured through dedicated devices, GPS receivers have become a common feature in smartphones and have been shown to produce comparable data to traditional devices [32-34]. Furthermore, GPS-enabled smartphones supplement traditional approaches with data from the cellular network to improve the speed with which location data are obtained [34].

Raw participant GPS records include latitude, longitude, accuracy, and the time at which the coordinate was recorded. These records alone are insufficient to draw inferences about participant behaviors and require post-processing to extract meaningful data on periods of movement, travel mode, and significant locations visited [29]. Geofencing and cluster-detection approaches are two commonly used approaches to infer visited locations from participant GPS data [31,43,49]. Geofencing requires that the researcher know the location for each feature of interest (e.g. health facility) where visits are to be detected. A boundary is then defined around each feature, which may take the form of the building's footprint or may be a radius about a specified point of interest. Participant GPS data are then examined for sequential points located within the specified boundary, classifying those points as a visit when the number of points or the duration between the first and last point meet some predefined threshold. Several cluster-based approaches exist for visit detection with substantial variability in their methods [50-53] One example, ST-DBSCAN, identifies clusters according to three parameters: a minimum number of points, and a maximum spatial and non-spatial (e.g. time) distance that these points are located from one another [54]. Cluster-based approaches are more flexible than geofencing approaches in that they do not require any knowledge of where features are located; however, they tend to require larger datasets and consequently increased computing power.

Both approaches require user-specified parameters, optimal values for which vary with study aim, setting, and the mechanics of the specific approach being applied. Overly conservative parameters will fail to detect true visits while overly permissive parameters will falsely classify non-visits as visits. Studies detecting visits to general locations have generally applied duration values of 20 to 30 minutes [31,41,55,56], though at least one study applied a threshold of 5 minutes [57]. In contrast, a study identifying hospitalizations set a higher duration threshold of 4 hours [43]. There is less convergence around an optimal distance threshold with even those studies using similar duration parameters applying distance values ranging from 10 to 200 meters [41,56]. While several studies evaluate the performance of their approach after the fact, the specification of parameter values is typically based on *a priori* assumptions of optimal thresholds. A notable exception to this, Theirry and colleagues evaluated the performance of six parameter sets on 750 simulated trajectories to determine the optimal values [57]. Such exercises require a calibrating dataset where the true value is known for each record, which may not be feasible in all contexts. Given the sensitivity of visit detection approaches to parameter specification, the value of undertaking such a calibration exercise should not be overlooked.

The present study was conducted within the context of a larger study comparing maternally-reported care-seeking behavior with measures of care-seeking behavior derived from participant GPS data [85]. This paper presents the development and implementation of a two-step approach to detect health facility visits within the context of a low to middle-income country, wherein: 1) optimal parameter values were identified through a calibration exercise involving field-worker simulated health facility visits and 2) these parameter values were applied to the prospective detection of participant health facility visits, the performance for which was assessed through monthly prompted recall surveys. Originally considered as an approach by which maternally-reported care-seeking would be validated, this was not pursued due to the circularity implicit in also including maternal recall via the prompted recall survey.

Methods

Study Site and Health Facility Census

We conducted a prospective cohort study in the 22 villages of the Vadu Health and Demographic Surveillance System, located in rural Pune district, Maharashtra state, India. Prior to study initiation we conducted a census of all locations where a mother might seek care for childhood illness, including private, public, formal, and informal providers. Field workers classified each provider by type, obtaining the location coordinates for each provider using a Garmin e-Trex device.

Participant Enrollment and Follow Up

Participants were mothers ages 15-49 with at least one child under five years of age, randomly sampled from the population register of the Vadu HDSS and enrolled during field worker home visits. Consenting participants were randomly assigned to either the primary study group ("phone group"), cross-sectional comparison group, or longitudinal comparison group. Participants in the phone group were provided with a GPS-enabled smartphone and visited monthly during the six-month study period. Comparison group participants were not provided with phones and were visited either once (cross-sectional group) or monthly (longitudinal group) over the study duration. These comparison groups were included to evaluate potential biases in reported care-seeking behavior resulting from providing participants with smartphones and following them up over time, the results of which are described elsewhere [84]. Given the aims of this analysis, the focus is restricted exclusively to participants in the phone group.

Phone group participants were provided with a Sony Xperia E4 Dual SIM phone and a SIM card unlocked for unlimited mobile data and in-network communication. Participants were asked to charge the phone daily, keep it on during the day, and maintain certain phone settings according to study specifications (e.g. mobile data on, location settings set to “high accuracy”). We asked that participants carry the phone whenever seeking care, though otherwise encouraged participants to use it as if it were their own. Participants could use the secondary SIM slot for their own SIM card, though this was not necessary for the phone to function. Field workers assisted with any participant queries regarding device use and study instructions when distributing the phones and at support visits conducted three and 11-14 days after enrollment, respectively.

A baseline questionnaire conducted at enrollment collected participant and household sociodemographic information, care-seeking preferences, attitudes toward phone-specific elements of the study protocol. During the same visit, field workers recorded participant household location using s Garmin e-Trex device. Participants were then visited monthly over six months and administered a questionnaire about recent childhood illness and any subsequent care-seeking. After completing this questionnaire, participants with one or more health facility visits detected through the GPS approach were administered a supplementary questionnaire on these visits (see below). Field workers assessed participant compliance at each follow-up visit and provided support whenever required.

Participant Mobility Tracking and Detection of Health Facility Visits

Participant smartphones were preinstalled with TrackCare, an application developed for this study to record the phone's location at one-minute intervals and transmit these data hourly to a central study server. Details of the application's development and implementation were previously described, including GPS data quality (mean observation time of 152 days [84% completeness], median accuracy of 12 meters), and participant compliance (79% overall) [115].

Participant GPS data were analyzed to detect potential health facility visits using a parameter-based geofencing approach. The approach defined a radius around each identified facility location, d_{max}, within which participant coordinates would be considered as within range of the specified facility. Sequential coordinates within range of the facility for longer than a minimum duration threshold, t_{min}, were classified as a facility visit. Recognizing the potentially noisy nature of GPS data, we allowed coordinates to temporarily appear outside the range of a health facility for an interval, t_{int}, provided that the coordinates subsequently reentered the facility range. Clusters of GPS data meeting these criteria were classified as potential health facility visits with the time associated with the earliest coordinate designated as the visit start time and the time associated with the last coordinate designated as the visit end time.

Optimal parameter values were determined through field-worker simulated health facility visits. A random sample of 15 facilities was selected from those identified during the facility census. Two field workers were given one smartphone each, preinstalled with TrackCare, and instructed to visit each location as if they were taking a child for care. Simulated visits lasted 10 minutes each, which was decided to be representative of the actual visit duration for most episodes of non-severe illness.

GPS data generated during the simulation exercise were analyzed according to 432 parameter combinations: d_{max} = 15-50 meters, t_{min} = 0-5 minutes, and t_{int} = 2-10 minutes. The optimal combination specified d_{max} = 25 meters, t_{min} = 3 minutes, and t_{int} = 8 minutes. This combination correctly identified visits to 14 of the 15 simulated visits, while incorrectly identifying visits at four facilities without a simulated visit from among the 117 eligible facilities not visited during the simulation (Sensitivity = 93%, Specificity 96%, AUC = 0.95).

We applied this approach to participant GPS data to prospectively detected health facility visits during study implementation. Before each scheduled follow-up visit, we analyzed the previous two weeks of participant GPS data to detect health facility visits. When one or more visits were detected, a list of detected visits was prepared and distributed to field workers. This list included the facility location, visit date, start time, and end time. If multiple visits were detected to the same location on a given day, one row was included on the list for each detected visit. Field workers reviewed these lists with participants at the end of each follow-up visit, asking participants to confirm whether each detected visit actually occurred and, if so, whether it was related to care-seeking for childhood illness. Participant responses were subsequently entered electronically into a list-specific Excel file.

Measures and Statistical Analysis

The overall performance of the GPS-based visit detection approach was assessed based on the proportion of all detected visits that were subsequently confirmed by participants during the prompted recall survey, regardless of visit purpose. Logistic regression models were applied to further explore the factors associated with the detection of false positive visits with each detected visit representing a single observation Analysis was restricted to visits detected during participant follow-up periods with at least 50% GPS data completeness. Sensitivity analysis considered alternative data completeness thresholds (no threshold, >75%, >90%).

We explored the association between visit confirmation status and various aspects of health facility density, modeled according to kernel density estimation (KDE), a geospatial analysis technique for exploring the distribution of features (e.g. health facilities) in space [116]. The approach overlays a cone-shaped probability density function on each point, the density of which decreases with distance from the center. The width of the kernel function depends on the user-defined bandwidth, which establishes the radius around the point within which the density function is contained. Where two density functions overlap (e.g. due to multiple features located near one another), the density value in the overlapping section is the sum of the individual functions. Selecting the appropriate bandwidth is key, with excessively low or high values leading to under- or over-smoothing of the data, both of which interfere with the identification of trends within the data. We evaluated seven bandwidths ranging from 100m to 2500m, classifying the resulting maps into quintiles of facility density [117]. Visual inspection identified 500m as the optimal bandwidth (Figure 6.1), as this value eliminated the noise present at lower values while preserving the important regions lost to over-smoothing at higher values (Supplementary Figures 6.1-6.7). Of the 232.2 km^2 included in the study area, Zone 1 accounts for 97% (224.7 km^2), Zone 2 accounts for 2% (4.6 km^2), Zone 3 accounts for 1% (1.7 km^2), and Zones 4 and 5 each account for less than 1% (0.8 and 0.5 km^2, respectively).

Explanatory variables included characteristics of the detected visit, participant location characteristics, and sociodemographic participant characteristics. Characteristics of the detected visit included duration, timing (overnight, other), proximity to participant residence and main highway, detected location sector (public, private) and type (hospital/clinic, other), location residence (urban, rural), and HF density zone at detected location (Zone 1, other). Participant location characteristics included residence (urban, rural), proximity to main highway, health facility density zone of residence, and average health facility density zone for GPS coordinates submitted during follow-up period. Participant sociodemographic characteristics included maternal age, education and employment status; previous household smartphone ownership; and household socioeconomic status (SES), defined according to the principal components analysis approach used by the Demographic and Health Survey [86].

Proximity to the main highway and health facility density zones at both the location and participant level were calculated using ArcGIS version 10.3 [88]. Average participant health facility density zone was calculated by plotting all participant GPS coordinates submitted during each follow-up period, identifying the corresponding health facility density zone for each coordinate, and computing the average value for the period. While the health facility density zone linked with each participant residence is likely to be highly correlated with the average zone during the follow-up period, the latter measure should provide a more accurate measurement of the participant's exposure to areas of high and low health facility density.

We estimated the unadjusted and adjusted associations between individual predictors and visit confirmation status through bivariate and multivariable logistic regression models, respectively, with standard errors adjusted for clustering among observations from the same participant. All variables demonstrating a marginally significant association with visit confirmation status in bivariate analysis ($p < 0.10$) were considered for inclusion in the multivariable model. The final set of variables included in the model was guided by the Akaike's information criterion and the Hosmer-Lemeshow goodness-of-fit test.

Ethical Considerations

All participants provided written consent prior to group assignment. Before obtaining consent, field workers informed participants that those assigned to the phone group would receive a smartphone and would be allowed to keep the device regardless of study completion. Various safeguards ensured the privacy of participant location data. Data stored on participant phones were saved in an encrypted database and were erased once they had been successfully transferred to a central study server. The study protocol was approved by the ethics committees of the University of Edinburgh and K.E.M. Hospital Research Centre, Pune (Study ID No. 1415).

Results

We enrolled 200 mothers with a total of 324 children under five from June to September 2015, baseline characteristics for whom are presented in Table 6.1, stratified by health facility density zone of participant residence. One participant withdrew from the study before completing the first follow-up visit and has been excluded from analysis. Average participant age was 25.3 years (standard deviation [SD] 3.3) with 10.9 completed years of schooling (SD 2.7). Current employment was reported by 28% of participants and was highest among participants located in density zone 1. Previous smartphone ownership was reported by 69% of participants. Two thirds of households resided in one of the four urban villages located along the main highway, while the remaining third were distributed among the 18 rural villages. The proportion of households located in urban villages is lowest among participants in density zone 1 and increases by zone. Median distance from a participant's residence to the main highway was 0.5 kilometers (interquartile range [IQR] 0.2-2.5) with a median of one health facility located within 500 meters (IQR 0-11). Participants in higher density zones were located nearer to the highway and had a greater number of health facilities nearby.

A total of 196 provider locations were identified, including seven public sector health facilities, 93 private hospitals and clinics, and 68 pharmacies. The same premises were shared by 29 providers, resulting in 167 unique provider locations (e.g. private hospital on first floor with a pharmacy on ground floor). Baseline location characteristics stratified by density zone are presented in Table 6.2. Locations were most commonly in private sector (81%) and either hospitals or clinics (59%), though this pattern was reversed among density zone 1 (94% public; 76% non-hospital/clinic) and less pronounced in density zone 2 (61% private; 52% hospital/clinic). Most facilities were located near the main highway (median distance 0.1 km; IQR 0.0-0.3) with facilities in zones 3-5 located nearer the highway than facilities in zones 1-2. A high degree of clustering of health facilities was observed overall with a median of 17 other facilities located within 500 meters of each facility (IQR 4-25). This was lowest in zone 1 (median 0, IQR 0-0) and increased proportionally with density zone until zone 5 (median 34, IQR 25-35).

A summary of participant follow-up visits is presented in Table 6.3. Field workers completed 1098 follow-up visits (8% loss to follow-up) with completion higher during earlier follow-up visits. Participant GPS data corresponding to each follow-up period were analyzed to detect health facility visits, identifying one or more visits during 793 follow-up visits (72%). Prompted recall surveys were completed during 694 of these visits (88%), constituting the evaluable sample for the remaining analysis. The proportion of participant follow-up visits where a prompted recall survey was indicated but was not completed increased during follow-up visits 5 and 6 due to server-related issues (Figure 2). While participant data were continually transferred to the central study server, a delay in relaying these data to the local study site resulted in field workers receiving the lists of GPS-detected visits after completion of the corresponding follow-up visits.

The visit detection algorithm identified a total of 22,251 possible health facilities visits across all study visits with participants confirming 440 (2%) of these visits as having occurred. The remaining visits represent false positives. Characteristics of each visit stratified by confirmation status are presented in Table 6.4. Median GPS-detected visit duration was 13 minutes (IQR 7-28) with no significant difference between detected visits that were subsequently confirmed and those that were not. Visits detected overnight accounted for 28% of all detected visits but only 12% of confirmed visits ($p < 0.001$). The median distance between the facility at which each visit was detected and the corresponding participant's household was 0.1 km (IQR 0.0-0.6), with confirmed visits tending to be further from the participants household than visits that were not confirmed (1.4 vs. 0.1km, $p < 0.001$). Hospitals and clinics accounted for 59% of all detected visits with a larger proportion of confirmed visits occurring at hospitals and clinics than non-confirmed visits (76% vs. 59%, $p < 0.001$). The performance of the visit detection algorithm varied by provider type with

79% of visits detected at the rural hospital confirmed (11 visits) and both visits detected at the primary health center were confirmed. In contrast, none of the 854 combined visits detected at any of the government sub-centers, the NGO hospital, or included shops were confirmed. Private sector facilities accounted for a greater proportion of confirmed visits than non-confirmed visits (95% vs. 91%, $p = 0.003$). Detected visits were primarily at facilities located in urban villages with no significant difference by confirmation status.

The results of the bivariate and multivariable logistic regression models are presented in Table 6.5. The final model included the distance between the facility at which the visit was detected and the participant's residence, the facility type (hospital/clinic, other), the density zone of the facility (zone 1, other), average participant density zone during follow-up, maternal employment status and follow-up visit number. False positives were more likely among visits detected nearby participant residences with the likelihood of detecting a false positive decreasing inversely with distance from a participant (odds ratio [OR] 0.89, 95% confidence interval [CI] 0.82-0.97). Visits detected at facilities other than hospitals and clinics were more likely to be false positives (OR 5.29, 95% CI 1.65-4.67) as were visits detected at facilities located in the zone of lowest health facility density (OR 5.29, 95% CI 1.74-16.05). The likelihood of detecting false positives also increased as participants spent more time in areas of increased health facility density (OR 2.29, 95% CI 1.62-3.25). Maternal employment status was associated with detected visits being false positives (OR 3.78; 95% CI 1.79-7.97) as was follow-up visit number (OR 1.19; 95% CI 1.02-1.39). The time of day when the visit was detected and the density zone within which participant households were located were significantly associated with confirmation status in bivariate analysis but were excluded from the final model. Visit timing was not significantly associated with confirmation status in the adjusted model and resulted in no improvement to model fit. In the case of household density zone, we compared models with household density zone and the average zone of participant GPS points during follow-up and found that the latter resulted in improved model fit.

Sensitivity analyses compared different cutoffs of GPS completeness and various specifications of the KDE bandwidth. Point estimates were consistent across all levels of GPS completeness, though confidence intervals for some associations that were significant at more permissive levels of data inclusion became non-significant as more data were excluded (Supplementary Figure 6.8). Point estimates and confidence intervals for facility type, maternal employment, and visit number were consistent between varying specifications of the KDE bandwidth (Supplementary Figure 6.9). Point estimates and confidence intervals for KDE-derived variables were highly sensitive to the bandwidth specification, with the modeled effect size of each variable attenuated as the bandwidth is increased.

Discussion

This study developed and implemented a parameter-based geofencing approach to detect health facility visits from passively collected participant GPS data. While similar approaches set parameters based on *a priori* assumptions of their optimal value, this study benefited from a calibration exercise where parameter values were determined using locally simulated health facility visits. The results of this calibration were then applied to the prospective detection of health facility visits among a cohort of 199 mothers with young children over six months of follow-up.

The overall performance of the GPS-based approach was low at 2%, measured as the proportion of all detected visits that were subsequently confirmed by participants during monthly prompted recall visits (440 confirmed of 22,251 detected).

Algorithms used to passively detect visited locations have struggled to balance detection of all true visits while minimizing the detection of false visits, though our results indicate lower performance than what has been reported elsewhere. Paz-Soldan and colleagues analyzed data collected from GPS trackers to detect visits to a variety of location types in a Peruvian city, with 47% of visits to health facilities and 41% of visits overall subsequently in participant interviews [31]. Nguyen and colleagues developed a GPS-based approach to detect hospitalizations across the United States, with participants confirming 65% of all detected hospitalizations [43]. In both cases a more conservative approach was applied to the classification of visits than was performed in our study. We applied three parameters to the detection of health facility visits: distance from the health facility (d_{max} = 25m), duration within that distance (t_{min} = 3mins), and interval during which points could temporarily appear outside that distance and still be considered part of the visit (t_{int} = 8mins). In contrast, the combination applied by Paz-Soldan and colleagues set d_{max} = 20m, t_{min} = 30mins, and t_{int} = 15mins [31]. The distance and duration threshold are more conservative than the values applied in our study, requiring that an individual be both nearer to a location and remain there over a longer period of time before being considered to have visited that location. Nguyen and colleagues specify an even higher duration threshold of 4 hours, though information is not provided on other parameters [43]. Three quarters of visits detected in our study had a duration less than 30 minutes with no difference between visits by confirmation status, so it is unclear whether applying a higher duration parameter would result in improved performance.

While our selected parameter combination performed well when applied to field worker-simulated health facility visits (AUC = 0.95), this combination resulted in a fifty-fold overestimation of health facility visits when applied to participant GPS data. This suggests a systematic difference between the simulation dataset and participant data collected throughout the study duration. Field workers were directed to visit several facilities over one day of testing, typically traveling directly from one facility to the next without spending time on other activities in the area surrounding the health facility. Field worker mobility during the simulation is therefore less likely to be representative of participants when moving through the study area. Consequently, field workers were less likely to engage in activities that might be associated with the detection of false positives (e.g. shopping at a location nearby a facility). The absence of such activities from the calibration dataset may have skewed the resulting parameter values. Also, the relatively short duration of each simulated visit (10 minutes) may have biased our duration value toward the value of three minutes, likely contributing to the high number of false positive detected. The combination of these factors may explain in part why the parameter combination performed well within the simulation exercise but poorly when applied to participant data. Future calibration efforts should include a broader range of simulated activities and make additional efforts to ensure that simulated data are representative of the participant behaviors under study. One recommendation for such an activity would be to enroll a small sample of participants, provide them with GPS-enabled smartphones, and ask that they document their travel behavior (e.g. through a travel diary). Such information could be used to train the detection algorithm to more accurately classify participant GPS data within the specific context of the study area.

Our analysis of confirmed visits provides valuable insight into the importance of study context. We identified five zones of health facility density with 40% of all health facilities located in the zone of highest density. This zone included two regions with a combined area of 0.5 km^2. Participants living in this zone represented only 4% of the study population yet accounted for 30% of all detected visits. Participant location was significantly associated with visit confirmation status for all measures examined in the unadjusted analysis. Visits were more likely to be false positives among participants residing in urban villages, nearer to the highway, and in areas of greater health facility density, measured either according to household location or the zone where participants spent time. The last of these covariates was highly associated with visit confirmation status in the adjusted model, where the likelihood that a detected visit was a false positive more than doubled with each one unit increase in the average zone where a participant spent her time. As a participant's exposure to highly urbanized areas with a high health facility density increases so too does the probability that she will pass within range of a health facility long enough for her data to indicate a visit to the facility. This complicates the implementation of a GPS-based approach for detecting visits to health facilities and other locations of interest in urban areas.

We also noted a significant association between health facility type and confirmation status. Visits detected at facilities other than hospitals and clinics were nearly three times as likely to be false positives as were visits at hospitals and clinics. Private sector facilities account for nearly all hospitals and clinics, while private pharmacies account for most other facility types. Within our study context, private hospitals and clinics are often located in similar areas. The nearly threefold reduction in performance among non-hospitals and clinics is therefore likely due to characteristics beyond facility location. Separately, we noted high performance at the government-run rural hospital (79%) and the primary health centre (100%), though only 16 visits were detected between the two. Both facilities are offset from the main road and are located in areas of lower density. While these limited numbers should be interpreted with caution, they suggest certain contexts where a similar approach may perform well.

Study Limitations

The study included several limitations. First, our study used a prompted recall survey to evaluate the performance of our visit detection algorithm. While this is a common approach for assessing the validity of GPS-based inferences, it involves a component of participant recall and may be subject to the same biases as survey-based methods [118]. Stopher and colleagues evaluated prompted recall within the context of a transportation study to identify travel mode and trip purpose, finding that prompted recall misclassified trip mode and trip purpose in about 10% and 20% of records, respectively [118]. The authors propose that wearable cameras may provide an alternative to prompted recall, though such an approach would not be feasible in our study for several reasons. Second, while our prompted recall approach allowed for the classification of detected health facility visits as true positives and false positives, it did not include any evaluation of instances where the GPS method identified no visit. This would have required the inclusion of an additional section where participants were asked to list any visits that occurred but were not among the GPS-detected visits. The inclusion of such a section was considered during study design but omitted due to concerns that this would even further prime participants to monitor their care-seeking behavior and unnecessarily bias the results of the parent study validating maternal care-seeking recall. Without confirmation of true negatives, the calculation of standard diagnostic metrics was not possible (e.g. sensitivity and specificity). Third, visit confirmation status was only collected for those visits detected according to the specified parameter combination, complicating the consideration of alternative parameter definitions after the fact. More conservative parameter values may result in fewer visits detected but evaluating its performance would only be possible when detected visits were a subset of those for which prompted recall data are available. Fourth, study implementation revealed instances where an improbable number of visits were detected during specific participant follow-up periods (e.g. >25 visits detected in previous two weeks). These lists were reviewed with participants according to the same protocol as all other lists, though their length may have resulted in non-standard completion of the prompted recall survey and potential outcome misclassification due either to participant fatigue or field worker improvisations. For example, a field worker may ask whether the participant visited any facility on a specific date rather than proceeding through each detected visit item in the list. Given our aim of exploring the performance of our visit detection algorithm, we have retained all visits in our analysis. Many factors association with these large lists were also of interest in our analysis (e.g. proximity between participant residence and health facility) and their exclusion would have biased our results. Finally, we assumed that GPS data collected from participant smartphones were a valid measurement of participant movement throughout the study period. This assumption held as long as the phone traveled with the participant but was violated when the participant traveled without the phone or when the phone was given to someone other than the participant. This is a common challenge faced by studies using GPS and other remote tracking technologies, though a recent study found the magnitude of misclassification resulting from participant non-compliance to be relatively low [107]. If a non-participant visited a health facility when carrying the phone, it is unlikely that any resulting GPS-based visit would be confirmed by the participant during prompted recall. While such instances would be rare, these would lower the calculated performance of the approach.

This study demonstrates the development and implementation of a GPS-based approach to detect health facility visits. Such approaches rely on externally defined parameter values, the selection of which is of critical importance. Differences in study settings provide no guarantee that the parameter combination applied in one setting will perform similarly in another. Future studies should carefully consider the process by which these parameters are set. There may also be interest in exploring how machine learning and other statistical techniques can be applied to refine a visit detection algorithm during study implementation. This study has also demonstrated the challenges associated with detecting health facility visits in areas of high density. Such an approach may perform much better in a more rural setting with fewer health facilities that are located further from urban centers. While we provided participants with smartphones, several studies have collected GPS data using participants' own smartphones [42,44]. As the discriminative capacity of visit detection algorithms increases, combining such an approach with data generated from participant-owned devices and publicly available spatial data could greatly expand the scale of current research into health-related mobility at relatively low cost.

Conclusion

While many studies have explored the capacity for GPS-based visit detection, few studies calibrate parameter values according to locally generated datasets. This study demonstrates the process by which locally simulated health facility visits informed the selection of optimal parameter values.

While these values demonstrated high diagnostic capacity when applied to the calibration dataset they resulted in a fifty-fold overestimation of visits when applied to participant-generated data, suggesting a systematic difference between the two datasets. While overall performance was low we observed several factors associated with performance. High clustering of health facilities within urban centers complicated the detection of health facility visits within this setting. Researchers interested in applying a GPS-based visit detection method within a context of similarly high health facility density should carefully consider the challenges posed by such a setting.

Funding

The study was funded by a prime award by the Bill and Melinda Gates Foundation (grant no. OPP1084442) made to the Johns Hopkins University of which University of Edinburgh, Edinburgh, UK and KEM Hospital Research Centre, Pune, India were sub-grantees.

Acknowledgements

Vadu HDSS is a member of the INDEPTH Network. We are grateful to the members of the Improving Coverage Measurement Core Group for their guidance in developing this study. We also thank Drs. Robert Black, Joanne Katz, Krishna Rao and Melinda Munos of the Johns Hopkins School of Public Health for their comments on the study protocol and this paper. Lastly, we wish to thank the study participants and the project staff at the Vadu HDSS, without whom this study would not have been possible.

Author Contributions

All authors conceived of the study and protocol. AM, PL, UC, TB, and SJ conducted data collection. AM analyzed the data and wrote the paper. SH, TB, HN, SJ, and HC provided guidance on the analysis and interpretation of results. All authors read and agree with the manuscript and conclusions.

Competing Interests

The authors declare no competing interests.

References

1 Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet (London, England). 2016;388:3027-35.

2 Jones G, Steketee RW, Black RE, Bhutta ZA, Morris SS. How many child deaths can we prevent this year? The Lancet. 2003;362:65-71.

29 Kerr J, Duncan S, Schipperijn J. Using global positioning systems in health research: a practical approach to data collection and processing. Am J Prev Med. 2011;41:532-40.

30 Enge P, Misra P. Special issue on Global Positioning System. Proceedings of the Ieee. 1999;87:3-15.

31 Paz-Soldan VA, Reiner RC, Jr., Morrison AC, Stoddard ST, Kitron U, Scott TW, et al. Strengths and weaknesses of Global Positioning System (GPS) data-loggers and semi-structured interviews for capturing fine-scale human mobility: findings from Iquitos, Peru. PLoS Negl Trop Dis. 2014;8:e2888.

32 Michael Y, McGregor EM, Allen J, Fickas S. Observing Outdoor Activity Using Global Positioning System-Enabled Cell Phones. In: Helal S, Mitra S, Wong J, Chang CK, Mokhtari M, editors. Smart Homes and Health Telematics: 6th International Conference, ICOST 2008 Ames, IA, USA, June 28-July 2, 2008 Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 177-84.

33 Wu J, Jiang C, Liu Z, Houston D, Jaimes G, McConnell R. Performances of different global positioning system devices for time-location tracking in air pollution epidemiological studies. Environ Health Insights. 2010;4:93-108.

34 Zandbergen PA, Barbeau SJ. Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones. Journal of Navigation. 2011;64:381-99.

41 Brusilovskiy E, Klein LA, Salzer MS. Using global positioning systems to study health-related mobility and participation. Social Science & Medicine. 2016;161:134-42.

42 Glasgow ML, Rudra CB, Yoo EH, Demirbas M, Merriman J, Nayak P, et al. Using smartphones to collect time-activity data for long-term personal-level air pollution exposure assessment. J Expo Sci Environ Epidemiol. 2016;26:356-64.

43 Nguyen KT, Olgin JE, Pletcher MJ, Ng M, Kaye L, Moturu S, et al. Smartphone-Based Geofencing to Ascertain Hospitalizations. Circulation Cardiovascular quality and outcomes. 2017;10.

44 Palmer JR, Espenshade TJ, Bartumeus F, Chung CY, Ozgencil NE, Li K. New approaches to human mobility: using mobile phones for demographic research. Demography. 2013;50:1105-28.

46 Krenn PJ, Titze S, Oja P, Jones A, Ogilvie D. Use of global positioning systems to study physical activity and the environment: a systematic review. Am J Prev Med. 2011;41:508-15.

49 Mazimpaka JD, Timpf S. Trajectory data mining: A review of methods and applications. Journal of Spatial Information Science. 2016;2016:61-99.

50 Xu R, Wunsch D, 2nd. Survey of clustering algorithms. IEEE Trans Neural Netw. 2005;16:645-78.

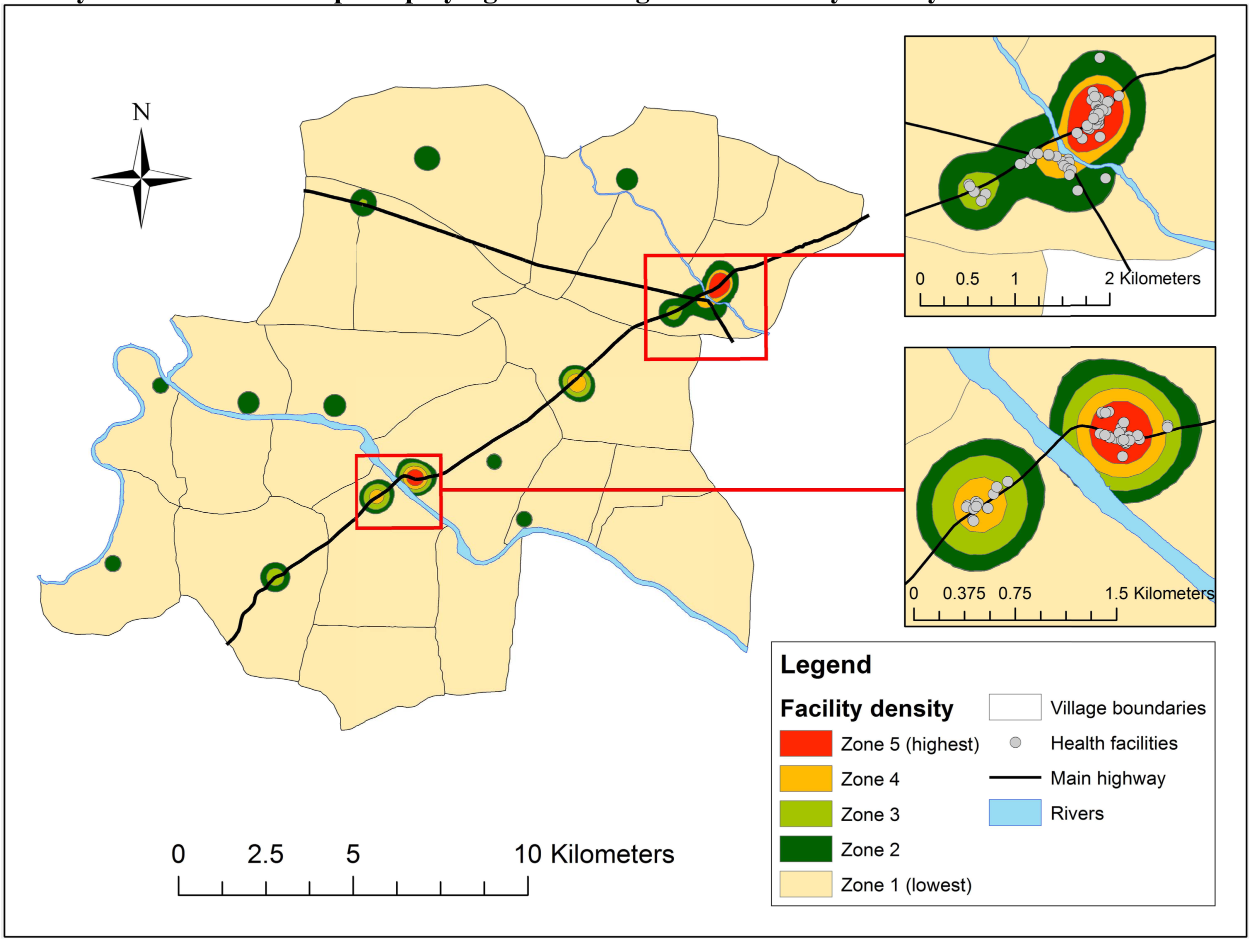
51 Palma AT, Bogorny V, Kuijpers B, Alvares LO. A clustering-based approach for discovering interesting places in trajectories. Proceedings of the 2008 ACM symposium

on Applied computing - SAC '08; Fortaleza, Ceara, Brazil. 1363886: ACM; 2008. p. 863-8.
52 Yuan G, Sun PH, Zhao J, Li DX, Wang CW. A review of moving object trajectory clustering algorithms. Artificial Intelligence Review. 2017;47:123-44.
53 Zheng Y. Trajectory Data Mining: An Overview. Acm Transactions on Intelligent Systems and Technology. 2015;6.
54 Birant D, Kut A. ST-DBSCAN: An algorithm for clustering spatial–temporal data. Data & Knowledge Engineering. 2007;60:208-21.
55 Donaire-Gonzalez D, Valentin A, de Nazelle A, Ambros A, Carrasco-Turigas G, Seto E, et al. Benefits of Mobile Phone Technology for Personal Environmental Monitoring. JMIR Mhealth Uhealth. 2016;4:e126.
56 Sanchez M, Ambros A, Salmon M, Bhogadi S, Wilson RT, Kinra S, et al. Predictors of Daily Mobility of Adults in Peri-Urban South India. International journal of environmental research and public health. 2017;14:783.
57 Thierry B, Chaix B, Kestens Y. Detecting activity locations from raw GPS data: a novel kernel-based algorithm. Int J Health Geogr. 2013;12:14.
75 MEASURE DHS. Demographic and Health Surveys. Available: http://www.measuredhs.com/. Accessed: May 16, 2018.
77 UNICEF. Multiple Indicator Cluster Survey (MICS). Available: http://www.unicef.org/statistics/index_24302.html Accessed: May 16, 2018.
84 Nair H, Williams L, Marsh A, Lele P, Bhattacharjee T, Chavan U, et al. Assessing the impact of mobile phones and repeated surveys on healthcare seeking behaviour for common childhood illnesses in rural India. 2018.
85 Hirve S, Marsh A, Lele P, Chavan U, Bhattacharjee T, Nair H, et al. Concordance between GPS-based smartphone app for continuous location tracking and mother's recall of care-seeking for child illness in India. 2018.
86 Filmer D, Pritchett LH. Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India. Demography. 2001;38:115-32.
88 ESRI. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute; 2011.
107 Isaacson M, Shoval N, Wahl HW, Oswald F, Auslander G. Compliance and data quality in GPS-based studies. Transportation. 2016;43:25-36.
108 Kroeger A. Health interview surveys in developing countries: a review of the methods and results. Int J Epidemiol. 1983;12:465-81.
109 Vazquez-Prokopec GM, Bisanzio D, Stoddard ST, Paz-Soldan V, Morrison AC, Elder JP, et al. Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. PLoS One. 2013;8:e58802.
110 Searle KM, Lubinda J, Hamapumbu H, Shields TM, Curriero FC, Smith DL, et al. Characterizing and quantifying human movement patterns using GPS data loggers in an area approaching malaria elimination in rural southern Zambia. R Soc Open Sci. 2017;4:170046.
111 Qi F, Du F. Tracking and visualization of space-time activities for a micro-scale flu transmission study. Int J Health Geogr. 2013;12:6.
112 Beekhuizen J, Kromhout H, Huss A, Vermeulen R. Performance of GPS-devices for environmental exposure assessment. J Expo Sci Environ Epidemiol. 2013;23:498-505.

113 Xu Y, Shaw SL, Zhao ZL, Yin L, Lu F, Chen J, et al. Another Tale of Two Cities: Understanding Human Activity Space Using Actively Tracked Cellphone Location Data. Annals of the American Association of Geographers. 2016;106:489-502.
114 Curriero FC, Pinchoff J, van Landingham SW, Ferrucci L, Friedman DS, Ramulu PY. Alteration of travel patterns with vision loss from glaucoma and macular degeneration. JAMA Ophthalmol. 2013;131:1420-6.
115 Marsh A, Bhattacharjee T, Hirve S, Lele P, Chavan U, Bhujbal S, et al. TrackCare: A smartphone-based approach for tracking participant movement in rural Pune district, India. 2018.
116 Chen Y-C A Tutorial on Kernel Density Estimation and Recent Advances. ArXiv e-prints [Internet]. Available: https://ui.adsabs.harvard.edu/#abs/2017arXiv170403924C. 'Accessed.'
117 King TL, Bentley RJ, Thornton LE, Kavanagh AM. Using kernel density estimation to understand the influence of neighbourhood destinations on BMI. BMJ Open. 2016;6.
118 Stopher PR, Shen L, Liu W, Ahmed A. The Challenge of Obtaining Ground Truth for GPS Processing. Transport Survey Methods: Embracing Behavioural and Technological Changes. 2015;11:206-17.

Tables and Figures

Figure 6.1 Map of Vadu Health and Demographic Surveillance System study area with health facility density zones and inset maps displaying areas of high health facility density



Note: Kernel density estimation based on 500-meter bandwidth. Zone 1 includes 97% (224.7 km^2) of the study area, Zone 2 includes 2% (4.6 km^2), Zone 3 includes 1% (1.7 km^2), and Zones 4 and 5 include less than 1% each (0.8 and 0.5 km^2, respectively).

Table 6.1 Baseline participant characteristics by health facility density zone

Characteristic	Total (N = 199) N (%)	Zone 1 (N = 123) N (%)	Zone 2 (N = 29) N (%)	Zone 3 (N = 19) N (%)	Zone 4 (N = 20) N (%)	Zone = 5 (N = 8) N (%)	P-value
Maternal age, mean (SD)	25.3 (3.3)	25.3 (3.3)	25.4 (3.4)	24.4 (3.9)	25.7 (2.7)	25.5 (2.3)	0.80
Maternal education (years), mean (SD)	10.9 (2.7)	10.9 (2.6)	10.5 (2.3)	11.2 (3.5)	10.6 (3.0)	11.9 (2.5)	0.71
Maternal employment	54 (28%)	43 (36%)	4 (14%)	3 (16%)	2 (10%)	2 (25%)	0.04
Wealth quintile							
1 (Lowest)	40 (20%)	18 (15%)	9 (31%)	6 (32%)	4 (20%)	3 (38%)	0.24
2	40 (20%)	23 (19%)	8 (28%)	4 (21%)	5 (25%)	0 (0%)	
3	40 (20%)	22 (18%)	7 (24.1%)	5 (26.3%)	4 (20%)	2 (25%)	
4	40 (20%)	30 (24%)	4 (13.8%)	1 (5.3%)	4 (20%)	1 (13%)	
5 (Highest)	39 (20%)	30 (24%)	1 (3.4%)	3 (15.8%)	3 (15%)	2 (25%)	
Previous smartphone ownership	132 (69%)	85 (71%)	13 (50%)	15 (79%)	13 (65%)	6 (75%)	0.21
Participant residence							
Rural	67 (34%)	61 (50%)	6 (21%)	0 (0%)	0 (0%)	0 (0%)	<0.001
Urban	132 (66%)	62 (50%)	23 (79%)	19 (100%)	20 (100%)	8 (100%)	
Distance to highway (km), median (IQR)	0.5 (0.2, 2.5)	1.6 (0.5, 3.1)	0.3 (0.2, 0.6)	0.2 (0.1, 0.3)	0.1 (0.0, 0.2)	0.1 (0.1, 0.2)	<0.001
Facility locations within 500m, median (IQR)	1 (0, 11)	0 (0, 1)	7 (6, 13)	13 (7, 18)	18 (14, 24)	32 (26, 40)	<0.001

Note: SD = standard deviation; km = Kilometer; IQR = Interquartile range; m = meter; Health facility density zones determined according to kernel density estimation using 500-meter bandwidth.

Table 6.2 Baseline facility characteristics by health facility density zone

Characteristic	Total (N = 167) N (%)	Zone 1 (N = 17) N (%)	Zone 2 (N = 23) N (%)	Zone 3 (N = 36) N (%)	Zone 4 (N = 34) N (%)	Zone = 5 (N = 67) N (%)	P-value
Sector							
Public	31 (19%)	16 (94%)	9 (39%)	3 (12%)	0 (0%)	3 (4%)	<0.001
Private	136 (81%)	1 (6%)	14 (61%)	23 (88%)	34 (100%)	64 (96%)	
Type, general							
Hospital/clinic	99 (59%)	4 (24%)	12 (52%)	17 (65%)	23 (68%)	43 (64%)	0.021
Other[1]	68 (41%)	13 (76%)	11 (48%)	9 (35%)	11 (32%)	24 (36%)	
Type, specific[2]							
Rural hospital	1 (<1%)	0 (0%)	1 (4%)	0 (0%)	0 (0%)	0 (0%)	<0.001
Primary health centre	1 (<1%)	1 (6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Sub-centre/ANM	5 (3%)	3 (18%)	1 (4%)	1 (4%)	0 (0%)	0 (0%)	
Anganwadi/ICDS centre	24 (14%)	12 (71%)	7 (30%)	2 (8%)	0 (0%)	3 (4%)	
NGO/trust hospital/clinic	1 (<1%)	0 (0%)	1 (4%)	0 (0%)	0 (0%)	0 (0%)	
Pvt. hospital	49 (29%)	0 (0%)	2 (9%)	10 (38%)	12 (35%)	25 (37%)	
Pvt. doctor/clinic	42 (25%)	0 (0%)	7 (30%)	6 (23%)	11 (32%)	18 (27%)	
Pharmacy/drugstore	40 (24%)	0 (0%)	1 (4%)	7 (27%)	11 (32%)	21 (31%)	
Shop	4 (2%)	1 (6%)	3 (13%)	0 (0%)	0 (0%)	0 (0%)	
Location							<0.001
Rural	34 (20%)	10 (59%)	19 (83%)	5 (19%)	0 (0%)	0 (0%)	
Urban	133 (80%)	7 (41%)	4 (17%)	21 (81%)	34 (100%)	67 (100%)	
Distance to highway (km), median (IQR)	0.1 (0.0, 0.3)	1.5 (0.9, 4.1)	2.7 (1.8, 3.8)	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	<0.001
Other locations within 500m, median (IQR)	17 (4, 25)	0 (0, 0)	2 (1, 3)	6 (5, 11)	17 (13, 18)	34 (25, 35)	<0.001

[1] Includes Anganwadi/ICDS centre, Pharmacy/drugstore, and Shop

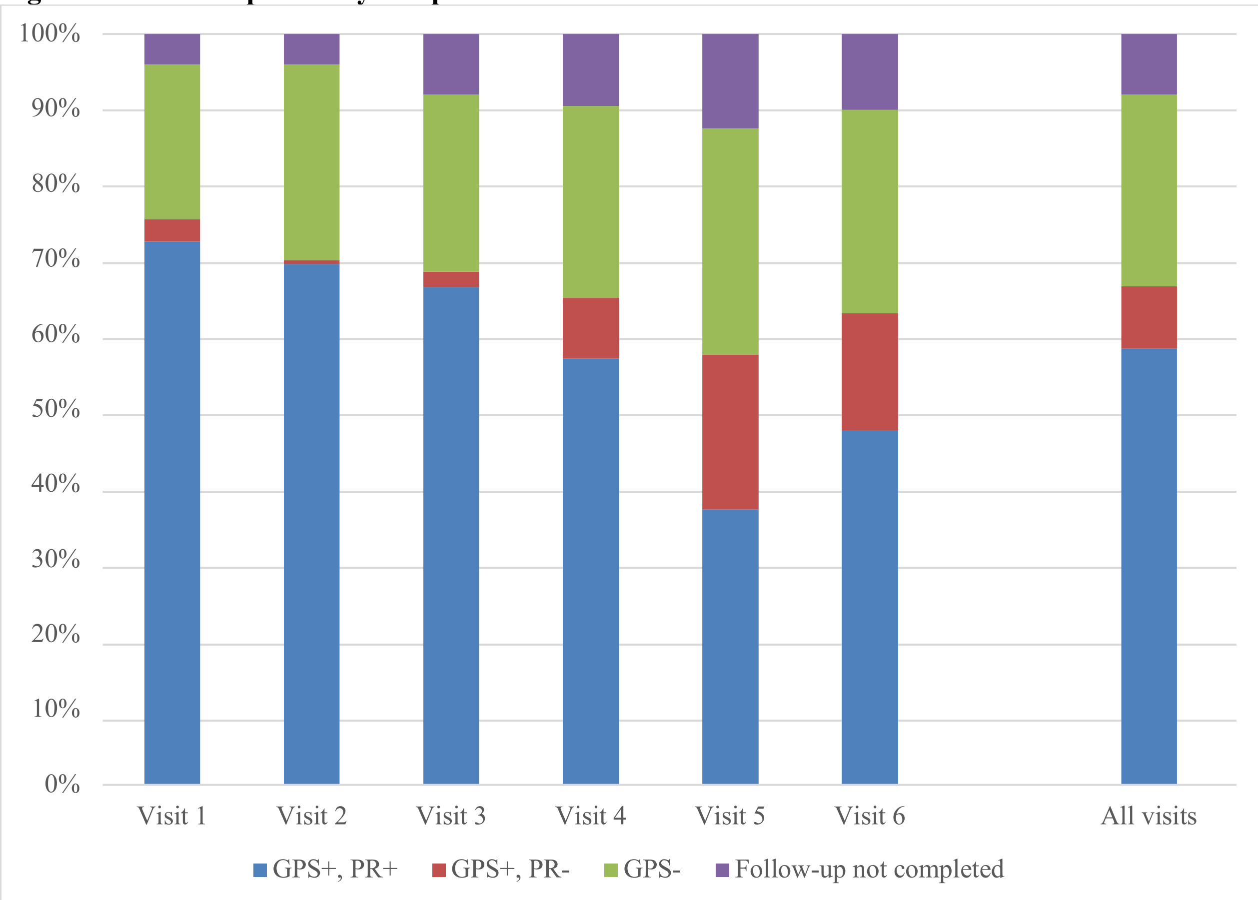
[2] In the case of 23 private hospitals and 5 private doctors/clinics, the same location coordinate was shared with a pharmacy/drugstore. The table presents the primary location type. For these shared location spaces, the primary type would be the private hospital or clinic.

Table 6.3 Follow-up visits by completion status

	(A) Follow-up visits planned	(B) Follow-up visits completed	(C) Follow-up visits with GPS visit indicated	(D) Follow-up visits with prompted recall completed	(E) Summary data from follow-up visits with prompted recall completed	
					Completeness of GPS data	GPS-detected visits per list
	N	N (% of A)	N (% of B)	N (% of C)	Median (IQR)	Median (IQR)
All visits	1,194	1,098 (92%)	793 (72%)	694 (88%)	78% (60-92%)	8 (3-27)
Visit 1	199	191 (96%)	150 (79%)	144 (96%)	80% (61-92%)	8 (4-17)
Visit 2	199	191 (96%)	139 (73%)	138 (99%)	80% (56-92%)	7 (3-27)
Visit 3	199	183 (92%)	136 (74%)	132 (97%)	78% (60-91%)	8 (3-30)
Visit 4	199	180 (90%)	129 (72%)	113 (88%)	70% (55-88%)	9 (4-22)
Visit 5	199	174 (87%)	114 (66%)	73 (64%)	78% (62-94%)	12 (4-31)
Visit 6	199	179 (90%)	125 (70%)	94 (75%)	74% (59-92%)	7 (3-26)

Note: IQR = Interquartile range

Figure 6.2 Follow-up visits by completion status



Note: GPS+ = At least one health facility visit detected in participant data; GPS- = No health facility visits detected in participant data. PR+ = Prompted recall survey completed for GPS-detected health facility visits; PR- = Prompted recall survey not completed for GPS-detected health facility visits.

Table 6.4 Characteristics of detected visits by confirmation status

Characteristic	All Detected Visits (N = 22,251) N (%)	True Positives (N = 440) N (%)	False Positives (N = 21,8111) N (%)	P-value
Duration (minutes), median (IQR)	13 (7, 28)	13 (7, 31)	13.1 (7, 28)	0.49
Visit occurred overnight[1]	6,159 (28%)	54 (12%)	6,105 (28%)	<0.001
Distance from participant residence (km), median (IQR)	0.1 (0.0, 0.6)	1.4 (0.4, 3.5)	0.1 (0.0, 0.5)	<0.001
Distance from highway (km), median (IQR)	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	0.1 (0.0, 0.2)	<0.001
Facility sector				
Public	2,071 (9%)	23 (5%)	2,048 (9%)	0.003
Private	20,180 (91%)	417 (95%)	19,763 (91%)	
Facility type, general				
Hospital/clinic	13,206 (59%)	336 (76%)	12,870 (59%)	<0.001
Other[2]	9,045 (41%)	104 (24%)	8,941 (41%)	
Facility type, specific				
Rural hospital	14 (<1%)	11 (3%)	3 (<1%)	<0.001
PHC	2 (<1%)	2 (<1%)	0 (0%)	
Sub-centre/ANM	229 (1%)	0 (0%)	229 (1%)	
Anganwadi/ICDS centre	1,826 (8%)	10 (2%)	1,816 (8%)	
NGO or trust hospital/clinic	302 (1%)	0 (0%)	302 (1%)	
Pvt. hospital	6,302 (28%)	277 (63%)	6,025 (28%)	
Pvt. doctor/clinic	6,357 (29%)	46 (10%)	6,311 (29%)	
Pharmacy/drugstore	6,896 (31%)	94 (21%)	6,802 (31%)	
Shop	323 (1%)	0 (0%)	323 (1%)	
Facility location				
Rural	2,055 (9%)	41 (9%)	2,014 (9%)	0.95
Urban	20,196 (91%)	399 (91%)	19,797 (91%)	
Health facility density zone				
1 (Lowest density)	1,605 (7%)	11 (3%)	1,594 (7%)	<0.001
2	1,229 (6%)	44 (10%)	1,185 (5%)	
3	1,968 (9%)	47 (11%)	1,921 (9%)	
4	5,648 (25%)	51 (12%)	5,597 (26%)	
5 (Highest density)	11,801 (53%)	287 (65%)	11,514 (53%)	

[1] Visit began after 21:00 and ended before 07:00.
[2] Includes Anganwadi/ICDS centre, Pharmacy/drugstore, and Shop
Note: IQR = interquartile range; PHC = Primary Health Centre.

Table 6.5 Unadjusted and adjusted associations with false positive health facility visit detection

Characteristic	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Characteristics of the detected visit		
Duration, minutes	1.00 (1.00 – 1.00)	-
Visit occurred overnight[1]	3.32 (1.55 - 7.13)**	-
Distance from participant residence, km	0.78 (0.71 - 0.85)**	0.89 (0.82 - 0.97)*
Distance from highway, km	1.09 (0.86 - 1.39)	-
Location is private sector	0.58 (0.23 - 1.42)	-
Location type		
Hospital/clinic	REF	REF
Other[2]	2.57 (1.44 - 4.57)**	2.78 (1.65 - 4.67)**
Location in density zone 1	3.23 (1.06 - 9.85)*	5.29 (1.74 - 16.05)**
Location in urban village	1.01 (0.38 - 2.7)	-
Participant location characteristics		-
Urban village	2.74 (1.16 - 6.48)*	-
Distance from highway, km	0.8 (0.64 - 0.99)*	-
Zone of participant residence		
1 (lowest density)	REF	-
2	1.88 (0.82 - 4.33)	-
3	8.83 (1.55 - 50.33)*	-
4	4.32 (1.29 - 14.47)*	-
5 (highest density)	11.04 (2.08 - 58.73)**	-
Average zone during follow-up	2.04 (1.51 - 2.75)**	2.29 (1.62 - 3.25)**
Participant sociodemographic characteristics		-
Maternal age, years	1.03 (0.91 - 1.18)	-
Maternal education, completed years	1.09 (0.92 - 1.3)	-
Maternal employment	2.61 (0.89 - 7.69)†	3.78 (1.79 - 7.97)**
Previous smartphone ownership	1 (0.36 - 2.81)	-
Household wealth quintile		-
1 (lowest)	REF	-
2	1.42 (0.4 - 5.04)	-
3	0.72 (0.19 - 2.72)	-
4	0.66 (0.18 - 2.47)	-
5 (highest)	1.45 (0.32 - 6.64)	-
Follow-up visit number	1.16 (0.97 - 1.38)	1.19 (1.02 - 1.39)*

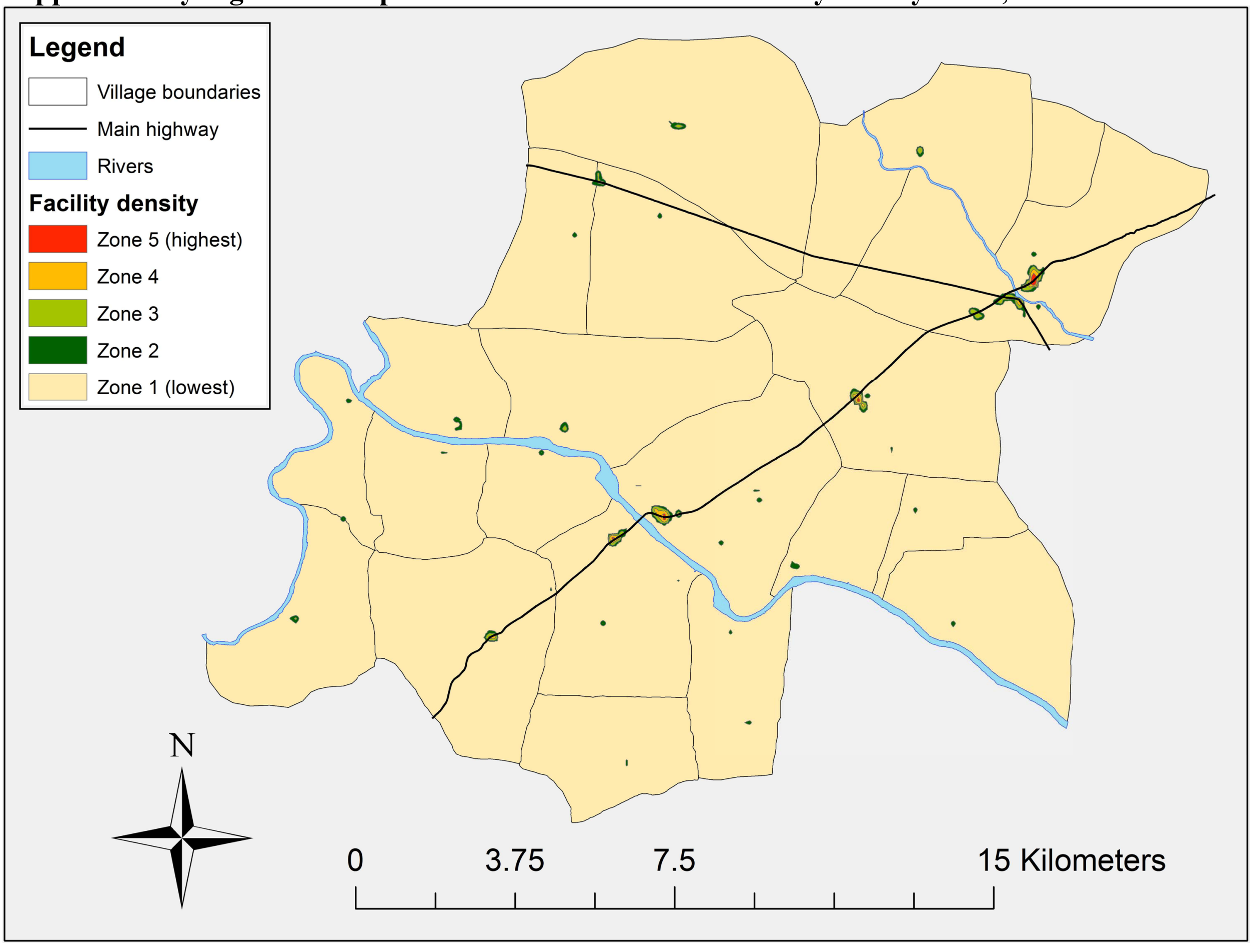
[1] Visit began after 21:00 and ended before 07:00.

[2] Includes Anganwadi/ICDS centre, Pharmacy/drugstore, and Shop

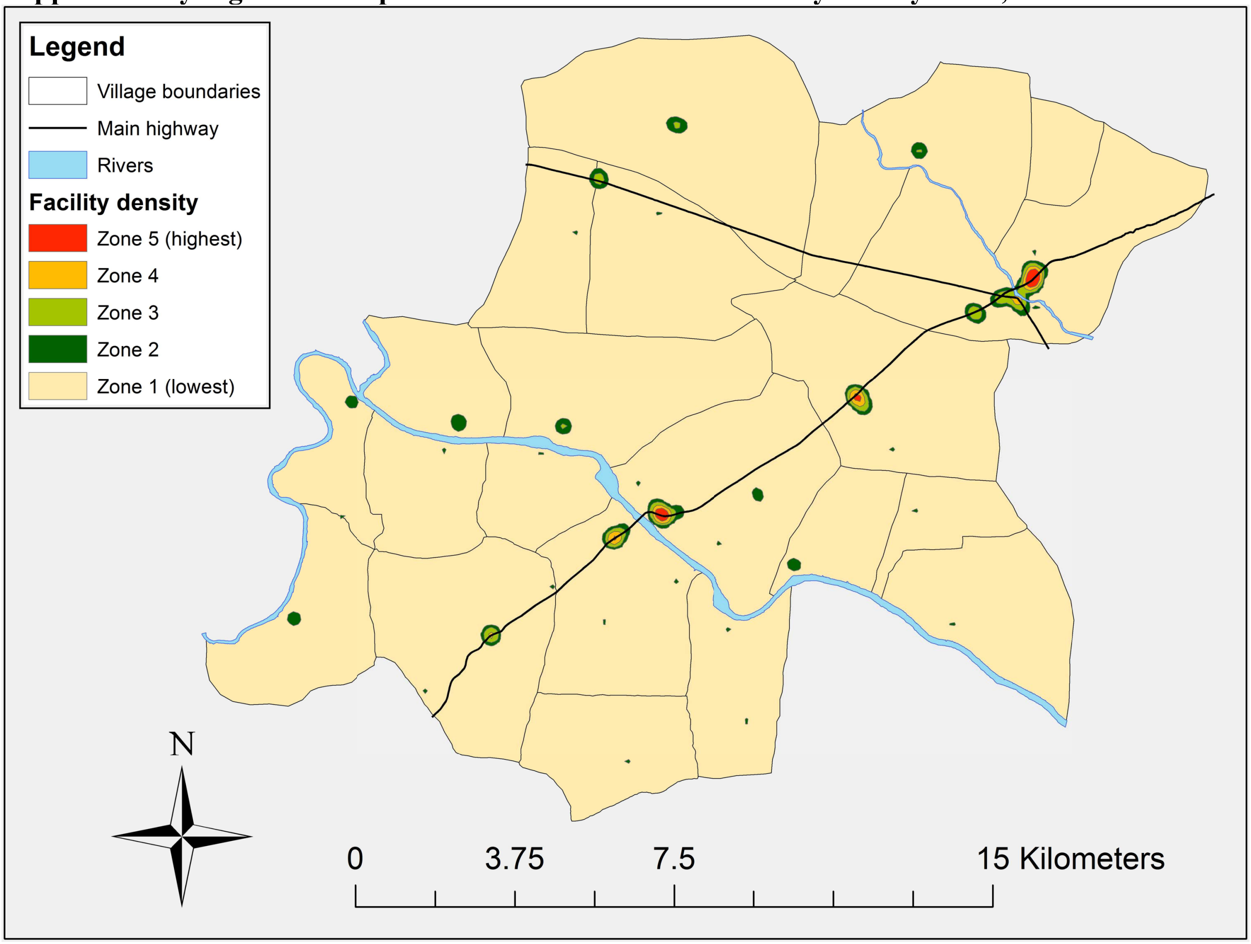
† $p < 0.10$; * $p < 0.05$; ** $p < 0.01$. Note: Confidence intervals adjusted for clustering among repeated observations from the same participant.

Supplementary Tables and Figures

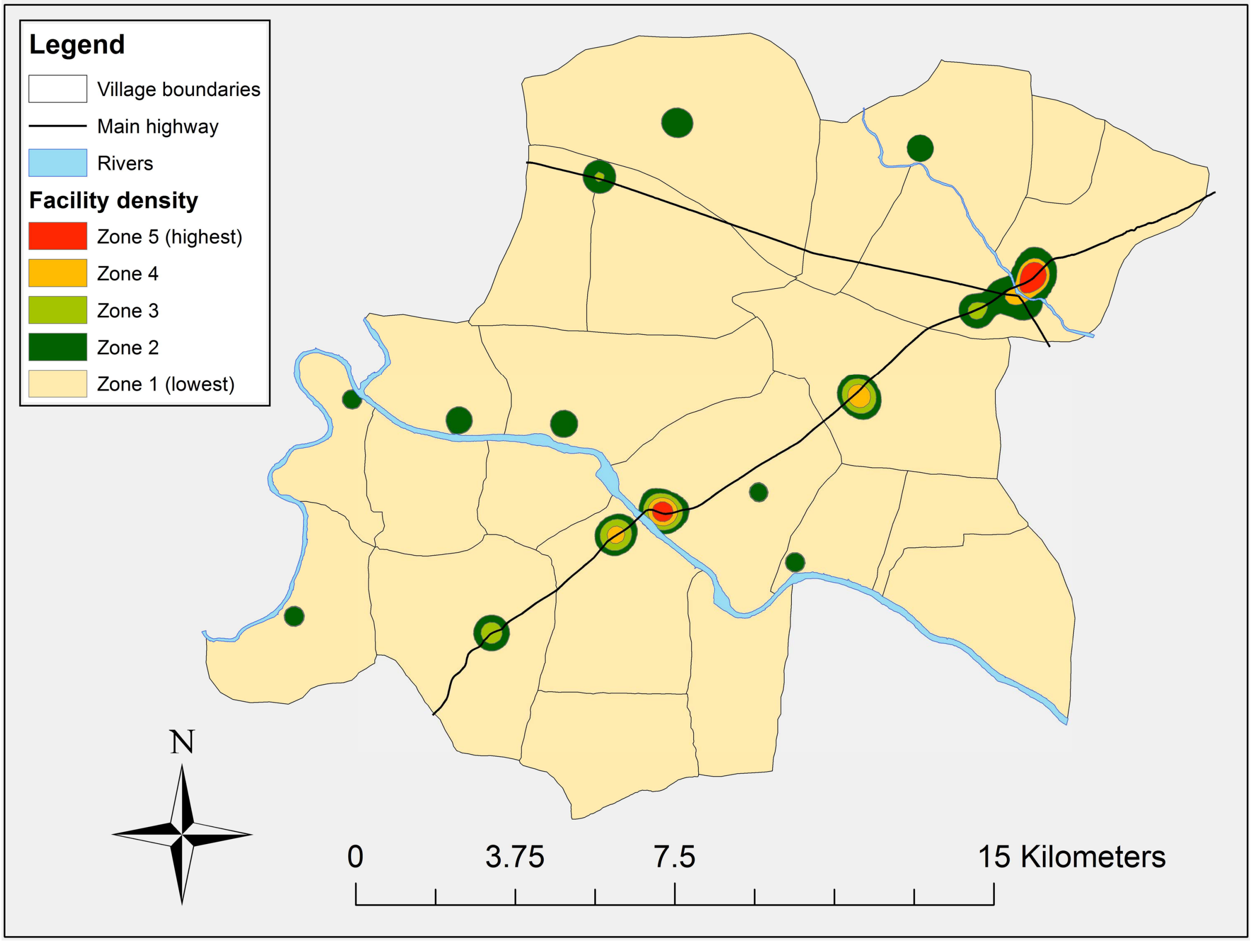
Supplementary Figure 6.1 Map of Vadu HDSS with health facility density zones, 100m bandwidth



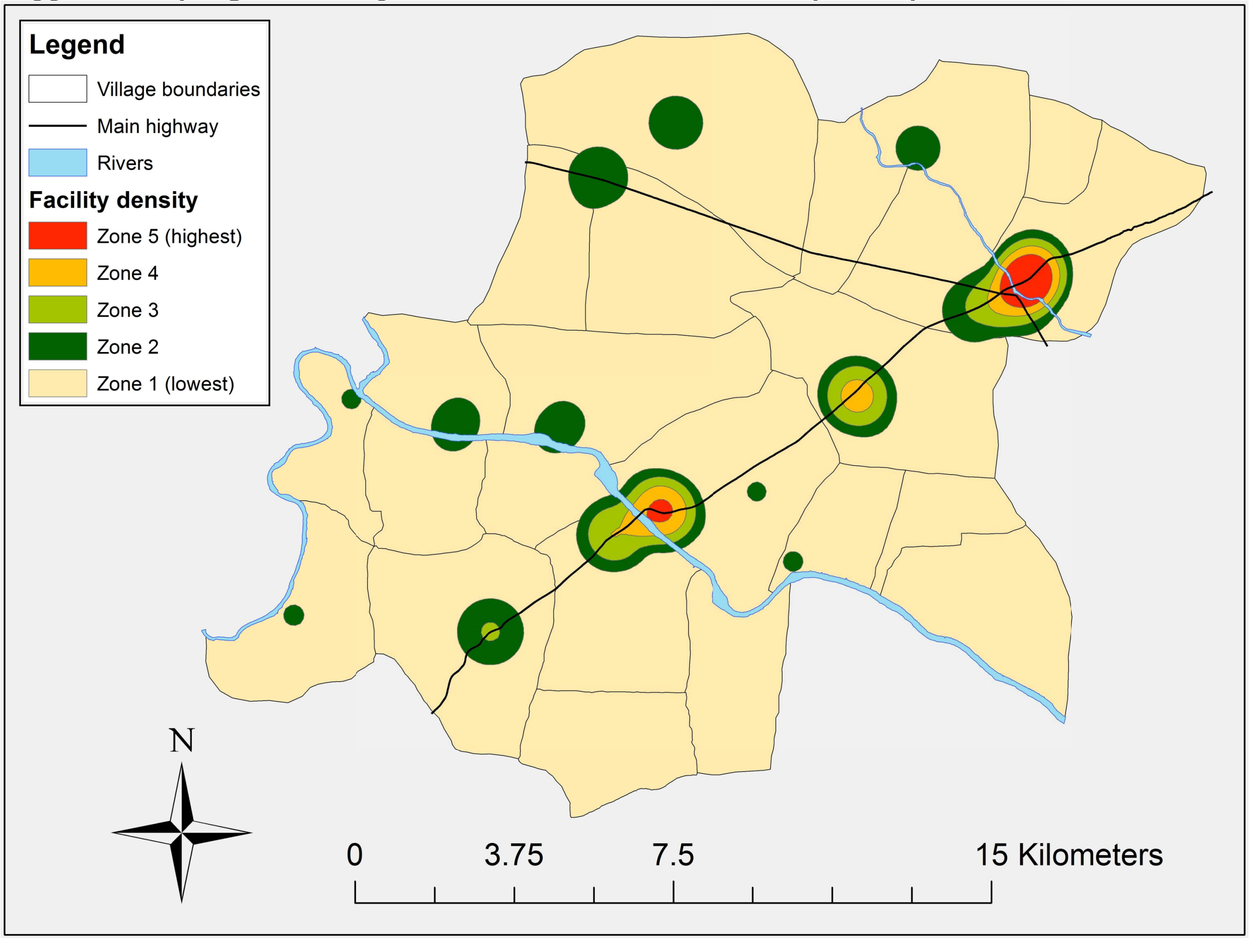
Supplementary Figure 6.2 Map of Vadu HDSS with health facility density zones, 250m bandwidth



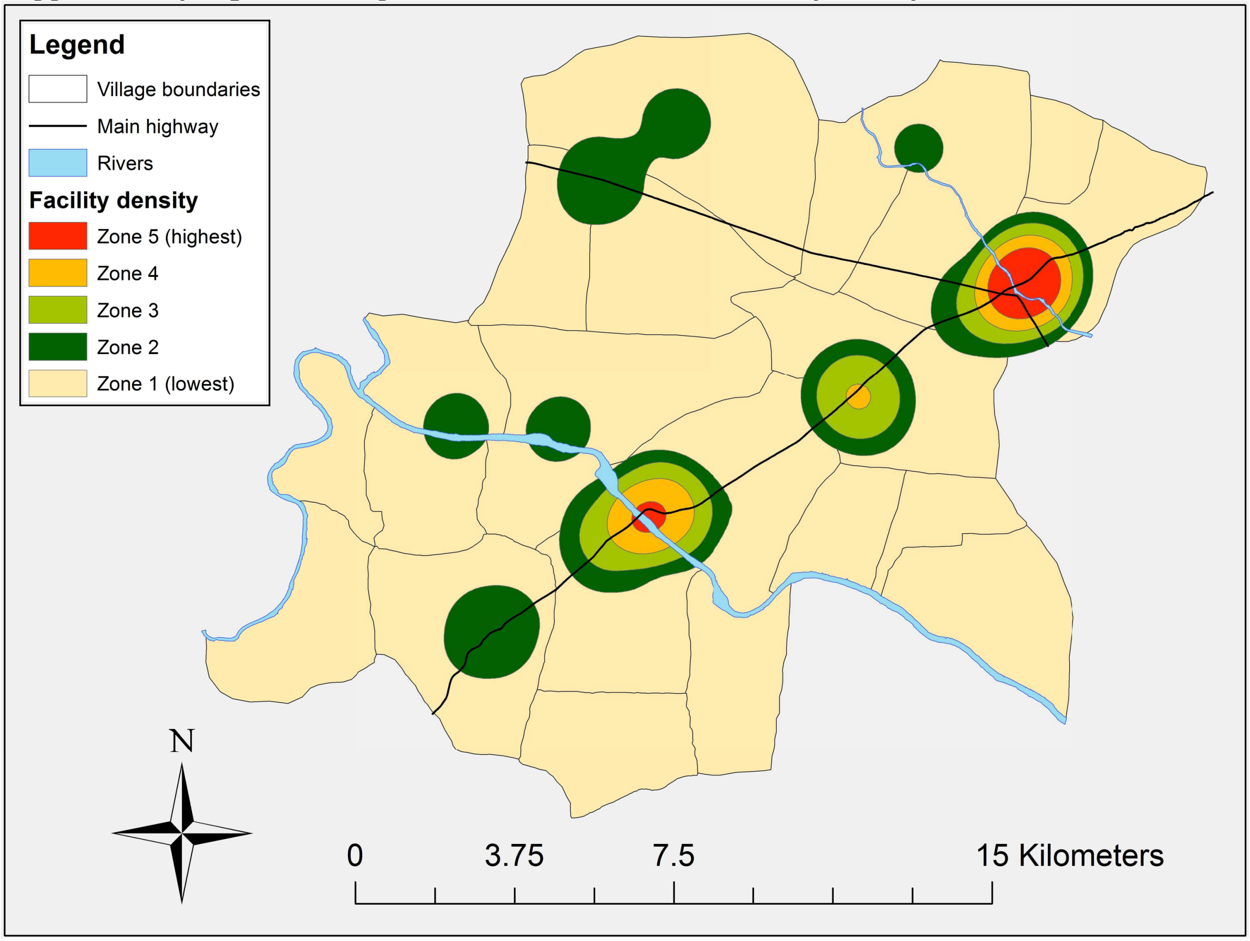
Supplementary Figure 6.3 Map of Vadu HDSS with health facility density zones, 500m bandwidth



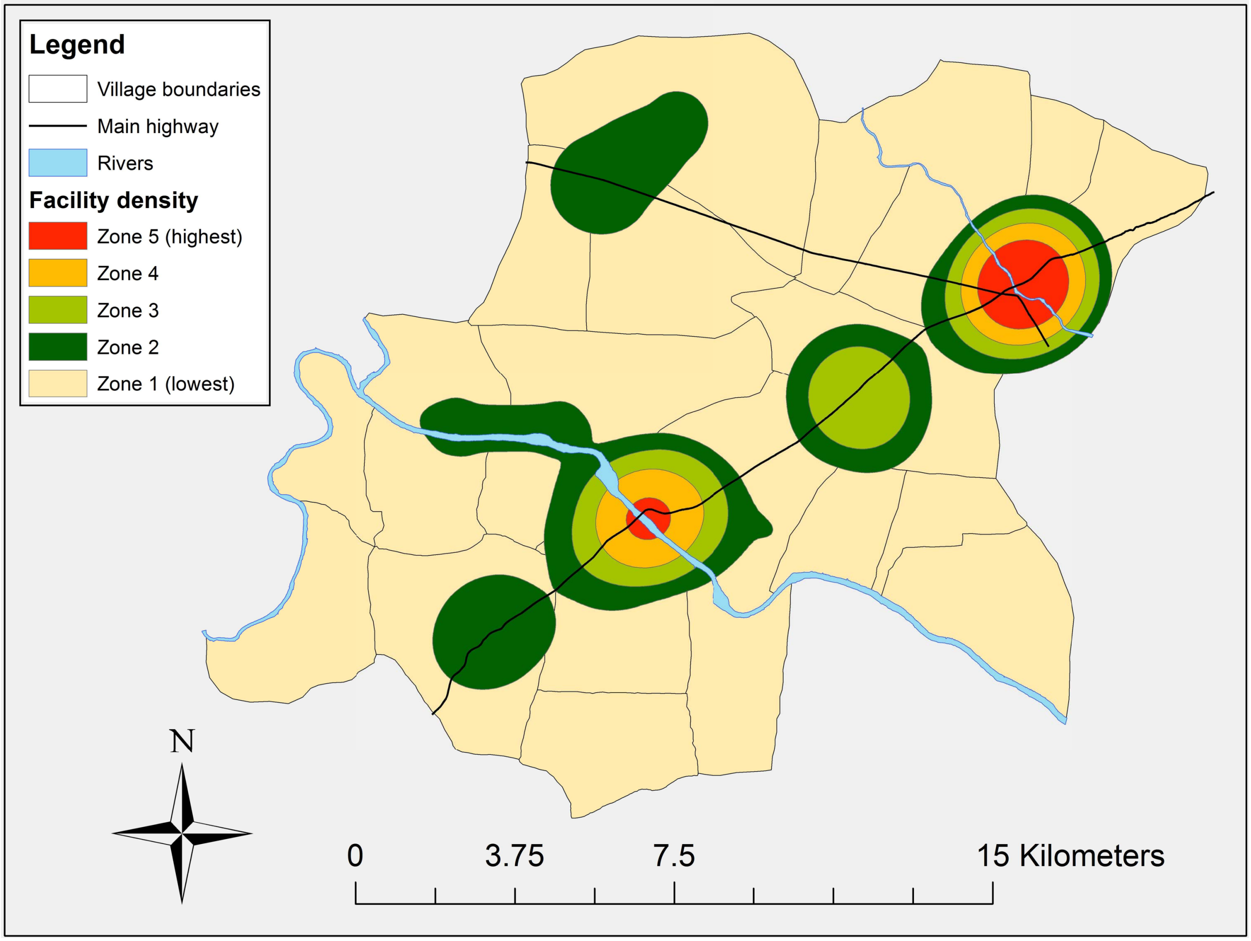
Supplementary Figure 6.4 Map of Vadu HDSS with health facility density zones, 1000m bandwidth



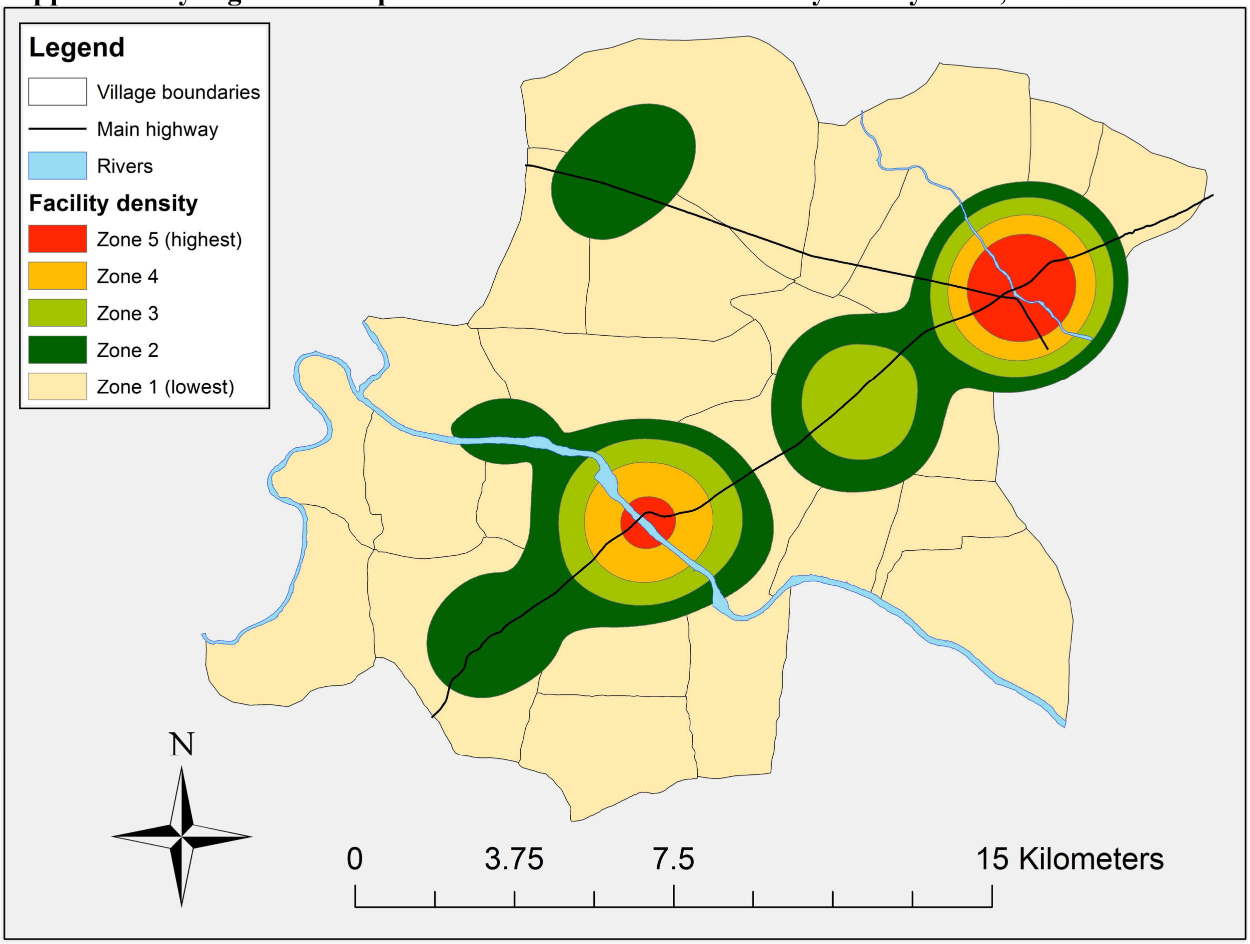
Supplementary Figure 6.5 Map of Vadu HDSS with health facility density zones, 1500m bandwidth



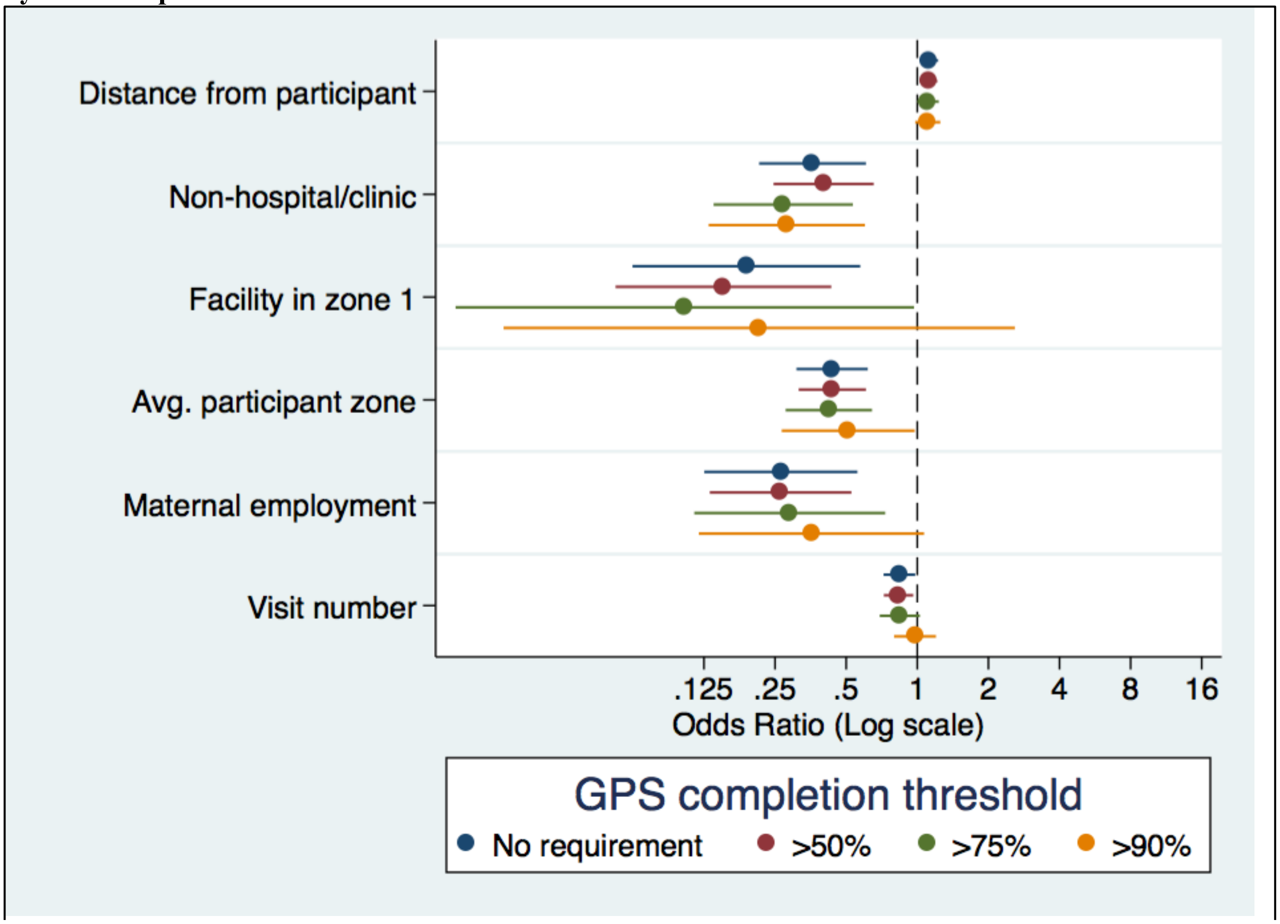
Supplementary Figure 6.6 Map of Vadu HDSS with health facility density zones, 2000m bandwidth



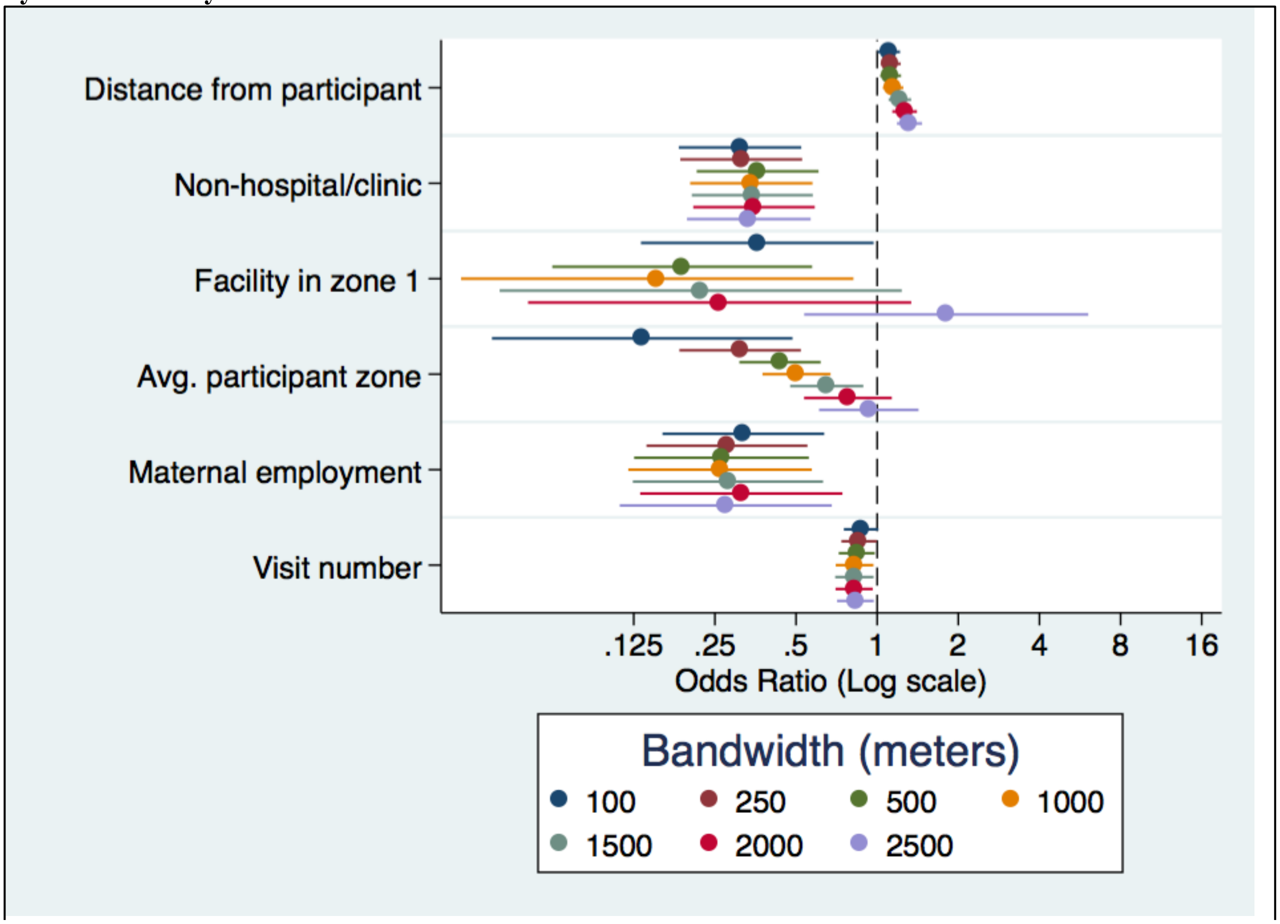
Supplementary Figure 6.7 Map of Vadu HDSS with health facility density zones, 2500m bandwidth



Supplementary Figure 6.8 Forest plot of multivariable logistic regression on visit confirmation status by GPS completion threshold used



Supplementary Figure 6.9 Forest plot of multivariable logistic regression on visit confirmation status by kernel density estimation bandwidth value



Note: Facility located in zone 1 perfectly predicted failure at a bandwidth to 250 meters and could not be estimated

Chapter 7: Discussion

7.1 Summary of Study Findings

Facility-based care-seeking for childhood illness is important for reducing mortality and morbidity associated with common childhood illnesses [2]. Large, nationally representative household surveys provide data on care-seeking practices, though the breadth of topics covered by such surveys limits the details available on any specific topics. Furthermore, the value of data collected through household surveys depends on the validity of participant responses. Drawing on a study comparing maternal recall of care-seeking for childhood illness with a GPS-based visit detection approach, the findings of this thesis research contribute to narrowing the gap in evidence on care-seeking practices in rural western India and provide evidence about the feasibility of alternative approaches to measure care-seeking behavior.

7.1.1 Objective 1: To evaluate the association between individual and household variables, including distance from a health facility, and the decision to seek care for an episode of childhood diarrhea, fever, or cough

An extensive body of research into the care-seeking practices in LMICs and specifically within India have revealed a set of core factors associated with care-seeking for childhood illness, including the type of care sought. The goal of this objective was to identify the set of determinants associated with care-seeking for childhood illness in rural Pune district, Maharashtra state, India. We also sought to describe the patterns of care-seeking for childhood illness, connecting each treatment step into care-seeking sequences. We hypothesized that the decision to seek care for an episode of childhood illness could be explained according to predisposing, enabling, and need characteristics as described by Andersen and subsequently adapted to the developing county context by Kroeger [4,8]. We observed high levels of care-seeking overall with illness characteristics as the key determinants of care-seeking. Children whose illnesses were reported as moderate-to-very severe were seven times more likely to be brought for care than children whose illness was non-severe and children whose mothers reported that they had more than one symptom were more than twice as likely to be brought for care. Self-medication through pharmacies constituted one tenth of reported care-seeking and was four times more likely to occur among illnesses perceived as non-severe compared to those perceived as more severe. Male bias in care-seeking has been documented within the study area and remains a concern in LMICs [69,119]. We observed no sex bias in overall care-seeking, though this does not rule out potential bias at other stages in the management of childhood illness. Maternal employment was inversely associated with care-seeking for illnesses perceived as non-severe. No association was observed between maternal employment and more severe illness. We hypothesize that maternal employment may increase the opportunity cost to mothers, the primary caregivers within our study setting, resulting in less care-seeking among illnesses perceived to be the least severe. We observed no association between care-seeking and maternal education, access to care, and household SES, though these null findings may be explained by relatively high levels of these variables within the study population.

7.1.2 Objective 2: To evaluate the feasibility of a smartphone-based approach for tracking participant movement and to explore factors associated with the approach's success

There is a growing body of evidence exploring the application of GPS technology to the study of health behaviors. GPS-enabled smartphones provide an attractive alternative to traditional devices with the potential for increased participant compliance and data completeness relative to traditional GPS devices. Various studies have applied GPS-enabled smartphones to the study of health behaviors though these studies are generally limited by small samples and short follow-up duration. Furthermore, few studies have applied this approach to care-seeking and none have done so within an LMIC context. The goal of this objective was to develop and implement a smartphone-based approach to track participant movement among a population of 200 mothers in rural Pune district, Maharashtra state, India. We were especially interested in data completeness, participant compliance with study procedures, and factors associated with both these variables. We hypothesized that the smartphone-based approach supported by monthly field-worker visits would result in increased data completeness and participant compliance relative to traditional approaches. We collected an average of 152 days of observation across 199 participants, constituting the largest study to collect movement data by smartphone and the first such study in a developing country to do so over an extended time period. Data completeness and participant compliance were both high. Data completeness was highest among participants complying with study procedures and residing in rural villages. Compliance was high overall and increased during the second half of the study. We also observed higher compliance among participants of increased socioeconomic status. We observed no association between compliance and maternal age, education, baseline perceptions, or previous smartphone ownership, suggesting that a smartphone-based approach may be appropriate among a broad range of participants.

7.1.3 Objective 3: To evaluate the correct detection rate of a GPS-based method to detect health facility visits and explore the factors associated with the improved performance of the method

While GPS has been widely applied to the study of health behaviors, few studies have applied a GPS-based approach to the detection of health facility visits. Various analytical approaches exist for inferring visited locations, including the detection of clusters of GPS points located near one another in space and time or identification of sequential GPS points within the proximity of some known location. These approaches are commonly based on various parameters, the values of which are often defined prior to study implementation and may not reflect the context within which the study is taking place or the nature of the activity being detected. The goal of this objective was to develop a GPS-based approach to detect health facility visits according to locally relevant parameter values and evaluate the approach's performance when applied to the detection of participant health facility visits. We hypothesized that the approach performance would vary with characteristics of the study participants, included health facilities, and their locations relative to one another. Through simulated health facility visits, we arrived at a parameter combination with both high sensitivity and specificity. While performance appeared high when applied to the simulated visits, we observed low overall performance when detecting health facility visits throughout the course of study implementation. The discrepancy between the approach's performance within the simulated dataset and during study implementation is likely due to systematic differences between the underlying data. Field workers moved directly from one facility to the next without engaging in the activities likely to generate false positives (e.g. visiting a location nearby a health facility). We observed that participants residing in areas of high health facility density and participants spending a high proportion of their time in such areas were more likely to generate false positive visits, highlighting the challenges of applying such an approach to the detection of health facility visits in an area of high health facility density. We hypothesize that this approach may perform better in areas where fewer facilities operate or where the facilities are located further from commercial centers. Some preliminary support for this is provided through examining two public facilities, the rural hospital and primary health center, both of which are offset from the main highway. While the total number of visits detected at these facilities is low, all visits detected at the primary health center were confirmed and nearly all detected at the rural hospital were confirmed

7.2 Implications for Policy and Practice

We observed high care-seeking overall, primarily from hospitals and clinics within the private sector. This is consistent with other studies within India, reporting low utilization of the public sector in favor of private sector providers. While private sector utilization accounts for the majority of care-seeking, there are concerns that these providers are poorly regulated and that the quality of services offered is variable [120,121]. Similarly, there is extensive cross-practice within the private sector in India, with providers trained in one system of medicine offering treatment from another system [122]. From the perspective of study participants, there is a clear preference toward private sector facilities. Combined with the relatively few public-sector providers operating within the study area, it would seem likely that the private sector will continue to provide for most of the medical needs of the study community in the foreseeable future. It is therefore important that measures be taken to ensure the quality of services provided at these providers. Separately, one should exercise caution when interpreting the absence of any male preference in care-seeking for childhood illness. Sex bias has been an important issue and may continue to be so for some time. We encourage future studies to consider whether sex bias is present at other stages in the management of childhood illness before forming a conclusion about the presence or absence of such bias within this community. Similarly, we caution against viewing maternal employment as an obstacle to care-seeking for non-severe illness. Maternal employment has been viewed as one component of increased maternal autonomy and its overall effect on child health and wellbeing may be positive. That we have observed a negative association with care-seeking for non-severe illness may have more to do with characteristics of the illness than with maternal employment.

We demonstrated the feasibility of a smartphone-based approach to prospectively track participant movement in a minimally invasive manner at large scale, both in terms of the number of enrolled participants and the duration of follow-up. High participant compliance with this approach suggests that similar, phone-based interventions may also hold promise within a population with high baseline phone ownership. It may also be possible to integrate similar approaches within participants' own phones to improve the scalability of such an approach. The raw participant data collected from this approach may be used for a variety of purposes and may be relevant to assess various environmental exposures. Our experience suggests that alternative approaches to monitor health behavior may be preferable for the immediate future within contexts with high health facility density. As the accuracy of these approaches improves, they may provide valuable information for improving coordination of care across various providers. Liss and colleagues recently reported on the findings of formative research into a smartphone-based approach to detect emergency room visits among high-risk patients and provide this information to patients' primary care providers

[58]. Similar approaches could be applied to monitoring antenatal care visits or compliance with referral, linking mothers with community health workers and nearby providers.

7.3 Strengths and Limitations

This thesis research benefitted from aspects of its design and implementation. First, the research was conducted within an established field site with a good working relationship in the community. This facilitated obtaining a sampling frame from which to select participants but also made it easier to follow participants over time. Second, the six-month study duration enabled us to consider each thesis objective across various points in time rather than at a single measurement. This provided us with greater insight into participant care-seeking practices than is available through the most common cross-sectional survey methods. This also enabled us to evaluate the feasibility of our smartphone-based approach over a much longer duration than is typically considered for comparable studies, finding that this approach appeared feasible up to at least six months. Third, using a modified DHS questionnaire allowed us to collect comparable data to what is currently collected while also probing further on care-seeking practices. The supplementary information obtained through these additional modules allowed us to construct care-seeking sequences that provide greater depth than is currently included within standard survey-based approaches. Fourth, the simulation of health facility visits prior to deploying our GPS-based visit detection algorithm provided our study with a set of locally derived parameters. While the overall performance of our visit detection approach was low, we gained valuable experience in how such parameters can be defined in a way that reflects local circumstances. Finally, the inclusion of a prompted recall component gave us the opportunity to explore factors associated with the performance of our visit detection algorithm. This enabled us to identify a subset of over 400 visits that were correctly identified, providing an opportunity to explore the circumstances where this approach performed best.

Our analysis of participant care-seeking behavior was subject to various limitations. First, data on childhood illness and related care-seeking behavior were collected during participant interviews and may be subject to a combination of recall and social desirability bias. Previous experience within this study area noted the potential for underreporting of sensitive health behaviors (e.g. abortion services), especially when solicited through a survey-based approach [100], though the same degree of bias is unlikely when asking about less sensitive behaviors. Second, we classified illness severity according to mothers' perception rather than through the presence of clinically-defined symptoms. The perception of severity has been associated with actual severity, though culturally-specific interpretations of signs and symptoms may also lead to concern [80]. We also expect that illness perception contributes more directly to a mother's decision to seek care than do the underlying clinical symptoms. Finally, the relatively high levels of sociodemographic indicators (e.g. mother's education, access to care) observed within our study population limit the generalizability of our findings to the state of Maharashtra and other contexts.

We also faced limitations in the collection and analysis of participant GPS data. First, the large volume of data generated each day by our smartphone-based approach limited our ability to respond to all identified issues in a timely fashion. Second, location data collected by the phone may also misclassify participants' true location during periods when the phone was not carried or when the phone was carried by someone else. This potential misclassification is common to GPS-based studies, though the overall impact may be small. If a non-participant visited a health facility when carrying the phone, it is unlikely that any resulting GPS-based visit would be confirmed by the participant during prompted recall. While such instances would be rare, these would lower the calculated performance of the approach. Third, compliance is measured according to a subset of phone-related behaviors that were readily observable at scheduled follow-up visits, though this measurement omits key components of compliance (e.g. charging behavior, lending the phone to others) and may not be representative of other time periods. Finally, while our prompted recall approach allowed for the classification of detected health facility visits as true positives and false positives, it did not include any evaluation of instances where the GPS method identified no visit. Visit confirmation status was only collected for those visits detected according to the specified parameter combination, complicating the consideration of alternative parameter definitions after the fact. More conservative parameter values may result in fewer visits detected but evaluating its performance would only be possible where detected visits were a subset of those for which prompted recall data are available.

7.4 Future Research and Next Steps

Our findings highlight several areas where further research could assist with understanding the questions we set out to investigate.

Within the context of high health facility utilization, it may be beneficial to go beyond exploring the determinants of visiting any health facility and explore the determinants associated with specific facility selection. This has been explored previously with respect to the concept of bypassing, providing valuable information on the specific household and provider-specific characteristics associated with provider selection [82]. Understanding bypassing behavior was beyond the scope of this thesis research but may be possible with currently available data. Separately, we classified care-seeking from a health facility based on facility type without accounting for the quality of care provided at individual facilities. Collecting data on quality of care during consultation for childhood illness was beyond the scope of this study. While we observed high levels of facility-based care-seeking overall, we expect that the quality of care received to vary according to participant and facility-specific characteristics. Future studies should consider investigating the quality of services that participants receive at these various providers to better understand the effective coverage of these services within the community level.

We noted that the perception of illness severity is highly associated with the decision to seek care. There is evidence to suggest that illness perception is driven in part by actual symptoms of the illness, though some culturally-specific factors may also determine the signs and symptoms associated with the perception of severe illness. A better understanding of local perceptions of illness severity would provide valuable insight into how well these perceptions reflect true illness severity. Should illness perceptions diverge from actual illness severity, this could be addressed through public health education campaigns.

Further research is needed to understand our null findings with regard to sex bias in care-seeking. This contrasts with previous findings in the study area and elsewhere in India. We have been cautious in our interpretation of this null finding due to the implications of such a finding among policymakers and public health practitioners. Additional studies should consider the potential for sex bias across various other aspects in the management of childhood illness before drawing a conclusion regarding the presence or absence of bias. Similarly, additional studies should consider the role of maternal employment in care-seeking for childhood illness. We hypothesized that maternal employment would enable care-seeking and consequently result in increased care-seeking among households where the mother was employed. We observed the opposite with regard to non-severe illness and no effect for more severe illness. It is possible that employed mothers face an increased opportunity cost when seeking care as this would represent lost wages. This would be consistent with our observation that care-seeking among employed mothers is lower for non-severe illness only. Additional studies are required to better understand the nature of this association.

We demonstrated the feasibility of collecting a large amount of participant GPS data, though various questions remain about this approach. First, would such an approach be equally as successful if deployed on participant-owned smartphones? We compared our results with two such studies and observed improved data quality relative to those studies. Pilot testing also revealed substantial heterogeneity in smartphone GPS performance. We hypothesize that the quality of data obtained from a study deploying such an approach on participant-owned devices would be highly associated with the quality of devices owned in that population. Within our study population many participants owned domestically manufactured devices, one of which was included in pilot testing and performed poorly compared to the two more expensive models manufactured abroad. Future studies should consider the quality of data collected through participant-owned devices in a variety of settings. Should these studies demonstrate comparable quality, this would greatly reduce the cost of implementing similar research projects in the future.

There is clearly much that remains to be explored with regard to approaches to detect health facility visits through participant GPS coordinates. This could begin with improving the manner in which the calibration exercise is conducted, enrolling pilot participants rather than using field workers to simulate health facility visits. Such an approach may yield a parameter set with improved diagnostic ability. There also exist several alternative methods to consider for the detection of health facility visits. These were outside the scope of this research as no prompted recall responses would be available. Future studies should consider whether these alternative methods, such as cluster detection, may provide improved inferences about visited locations. There may also be the potential for improving the diagnostic capacity for any approach over the course of the study, using participant responses during one round of visit identification to better train the visit detection algorithm for subsequent rounds. The computation requirements for such an approach should not be underestimated as even our less sophisticated approach proved much more computationally intensive than anticipated. As approaches to remotely detect health facility visits improve, these can be integrated with other research instruments. For instance, the detection of a health facility visit may trigger a short survey to appear, allowing participants to provide information on the reason for their visit, the quality of services, and other information. Such an approach is being investigated by Liss and colleagues, who are applying a smartphone-based approach to detect emergency room visits among high-risk patients and link this information with patients' primary care providers for follow up [58].

References

2 Jones G, Steketee RW, Black RE, Bhutta ZA, Morris SS. How many child deaths can we prevent this year? The Lancet. 2003;362:65-71.

4 Andersen RM. Families' Use of Health Services: A Behavioral Model of Predisposing, Enabling, and Need Components. Ann Arbor: Purdue University; 1968.

8 Kroeger A. Anthropological and socio-medical health care research in developing countries. Soc Sci Med. 1983;17:147-61.

58 Liss DT, Serrano E, Wakeman J, Nowicki C, Buchanan DR, Cesan A, et al. "The Doctor Needs to Know": Acceptability of Smartphone Location Tracking for Care Coordination. JMIR Mhealth Uhealth. 2018;6:e112.

69 Ganatra B, Hirve S. Male bias in health care utilization for under-fives in a rural community in western India. Bull World Health Organ. 1994;72:101-4.

80 Yoder PS, Hornik RC. Symptoms and perceived severity of illness as predictive of treatment for diarrhea in six Asian and African sites. Social Science & Medicine. 1996;43:429-39.

82 Akin JS, Hutchinson P. Health-care facility choice and the phenomenon of bypassing. Health Policy Plan. 1999;14:135-51.

100 Ganatra B, Hirve S. A community-based study on induced abortions. Issues Med Ethics. 2001;9:7-8.

119 Calu Costa J, Wehrmeister FC, Barros AJ, Victora CG. Gender bias in careseeking practices in 57 low- and middle-income countries. J Glob Health. 2017;7:010418.

120 Sengupta A, Nundy S. The private health sector in India. BMJ (Clinical research ed). 2005;331:1157-8.

121 Leena V Gangolli RD, Abhay Shukla. Review of Healthcare in India. Mumbai: Centre for Enquiry into Health and Allied Themes; 2005.

122 Math SB, Moirangthem S, Kumar CN. Public health perspectives in cross-system practice: past, present and future. Indian journal of medical ethics. 2015;12.

COMPREHENSIVE BIBLIOGRAPHY

1 Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, et al. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet (London, England). 2016;388:3027-35.

2 Jones G, Steketee RW, Black RE, Bhutta ZA, Morris SS. How many child deaths can we prevent this year? The Lancet. 2003;362:65-71.

3 Waldman RJ, Bartlett AV, Campbell CC, Steketee RW. Overcoming remaining barriers: the pathway to survival. Arlington, VA: BASICS Project, 1996.

4 Andersen RM. Families' Use of Health Services: A Behavioral Model of Predisposing, Enabling, and Need Components. Ann Arbor: Purdue University; 1968.

5 Andersen R, Newman JF. Societal and individual determinants of medical care utilization in the United States. The Milbank Memorial Fund quarterly Health and society. 1973;51:95-124.

6 Aday LA, Andersen R. A framework for the study of access to medical care. Health services research. 1974;9:208-20.

7 Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? Journal of health and social behavior. 1995;36:1-10.

8 Kroeger A. Anthropological and socio-medical health care research in developing countries. Soc Sci Med. 1983;17:147-61.

9 Bennett A, Eisele T, Keating J, Yukich J. Global trends in care seeking and access to diagnosis and treatment of childhood illnesses. Rockville, Maryland, USA: ICF International, 2015.

10 Geldsetzer P, Williams TC, Kirolos A, Mitchell S, Ratcliffe LA, Kohli-Lynch MK, et al. The recognition of and care seeking behaviour for childhood illness in developing countries: a systematic review. PLoS One. 2014;9:e93427.

11 Higgs G. A Literature Review of the Use of GIS-Based Measures of Access to Health Care Services. Health Services and Outcomes Research Methodology. 2004;5:119-39.

12 Guagliardo MF. Spatial accessibility of primary care: concepts, methods and challenges. Int J Health Geogr. 2004;3:3.

13 Nesbitt RC, Gabrysch S, Laub A, Soremekun S, Manu A, Kirkwood BR, et al. Methods to measure potential spatial access to delivery care in low- and middle-income countries: a case study in rural Ghana. Int J Health Geogr. 2014;13:25.

14 Costa LS, Nassi CD, Pinheiro RS, Almeida RM. Accessibility of selected hospitals and medical procedures by means of aerial and transit network-based measures. Health services management research. 2003;16:136-40.

15 Noor AM, Amin AA, Gething PW, Atkinson PM, Hay SI, Snow RW. Modelling distances travelled to government health services in Kenya. Trop Med Int Health. 2006;11:188-96.

16 Okwaraji YB, Cousens S, Berhane Y, Mulholland K, Edmond K. Effect of geographical access to health facilities on child mortality in rural Ethiopia: a community based cross sectional study. PLoS One. 2012;7:e33564.

17 Baume C, Helitzer D, Kachur SP. Patterns of care for childhood malaria in Zambia. Soc Sci Med. 2000;51:1491-503.
18 Bigogo G, Audi A, Aura B, Aol G, Breiman RF, Feikin DR. Health-seeking patterns among participants of population-based morbidity surveillance in rural western Kenya: implications for calculating disease rates. Int J Infect Dis. 2010;14:e967-73.
19 Wilson SE, Ouedraogo CT, Prince L, Ouedraogo A, Hess SY, Rouamba N, et al. Caregiver recognition of childhood diarrhea, care seeking behaviors and home treatment practices in rural Burkina Faso: a cross-sectional survey. PLoS One. 2012;7:e33273.
20 Glik DC, Ward WB, Gordon A, Haba F. Malaria treatment practices among mothers in Guinea. Journal of health and social behavior. 1989;30:421-35.
21 Gombojav N, Manaseki-Holland S, Pollock J, Henderson AJ. The effects of social variables on symptom recognition and medical care seeking behaviour for acute respiratory infections in infants in urban Mongolia. Arch Dis Child. 2009;94:849-54.
22 Das A, Ravindran TS. Factors affecting treatment-seeking for febrile illness in a malaria endemic block in Boudh district, Orissa, India: policy implications for malaria control. Malar J. 2010;9:377.
23 Chaturvedi HK, Mahanta J, Pandey A. Treatment-seeking for febrile illness in north-east India: an epidemiological study in the malaria endemic zone. Malar J. 2009;8:301.
24 Gao W, Dang S, Yan H, Wang D. Care-seeking pattern for diarrhea among children under 36 months old in rural western China. PLoS One. 2012;7:e43103.
25 Feikin DR, Nguyen LM, Adazu K, Ombok M, Audi A, Slutsker L, et al. The impact of distance of residence from a peripheral health facility on pediatric health utilisation in rural western Kenya. Trop Med Int Health. 2009;14:54-61.
26 Moisi JC, Kabuka J, Mitingi D, Levine OS, Scott JA. Spatial and socio-demographic predictors of time-to-immunization in a rural area in Kenya: Is equity attainable? Vaccine. 2010;28:5725-30.
27 Kroeger A. Response errors and other problems of health interview surveys in developing countries. World health statistics quarterly Rapport trimestriel de statistiques sanitaires mondiales. 1985;38:15-37.
28 Maas J, Sterkenburg RP, de Vries SI, Pierik F. Using GPS to Measure the Interaction Between Individuals and Their Neighbourhood. Neighbourhood Structure and Health Promotion2013. p. 153-75.
29 Kerr J, Duncan S, Schipperijn J. Using global positioning systems in health research: a practical approach to data collection and processing. Am J Prev Med. 2011;41:532-40.
30 Enge P, Misra P. Special issue on Global Positioning System. Proceedings of the Ieee. 1999;87:3-15.
31 Paz-Soldan VA, Reiner RC, Jr., Morrison AC, Stoddard ST, Kitron U, Scott TW, et al. Strengths and weaknesses of Global Positioning System (GPS) data-loggers and semi-structured interviews for capturing fine-scale human mobility: findings from Iquitos, Peru. PLoS Negl Trop Dis. 2014;8:e2888.
32 Michael Y, McGregor EM, Allen J, Fickas S. Observing Outdoor Activity Using Global Positioning System-Enabled Cell Phones. In: Helal S, Mitra S, Wong J, Chang CK, Mokhtari M, editors. Smart Homes and Health Telematics: 6th International Conference, ICOST 2008 Ames, IA, USA, June 28-July 2, 2008 Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 177-84.

33 Wu J, Jiang C, Liu Z, Houston D, Jaimes G, McConnell R. Performances of different global positioning system devices for time-location tracking in air pollution epidemiological studies. Environ Health Insights. 2010;4:93-108.
34 Zandbergen PA, Barbeau SJ. Positional Accuracy of Assisted GPS Data from High-Sensitivity GPS-enabled Mobile Phones. Journal of Navigation. 2011;64:381-99.
35 Benson AC, Bruce L, Gordon BA. Reliability and validity of a GPS-enabled iPhone "app" to measure physical activity. Journal of sports sciences. 2015;33:1421-8.
36 Garnett R, Stewart R. Comparison of GPS units and mobile Apple GPS capabilities in an urban landscape. Cartography and Geographic Information Science. 2015;42:1-8.
37 Jahkola O, Happonen A, Knutas A, Ikonen J, editors. What should application developers understand about mobile phone position data. ACM International Conference Proceeding Series; 2017.
38 Wu LL, Yang B, Jing P. Travel Mode Detection Based on GPS Raw Data Collected by Smartphones: A Systematic Review of the Existing Methodologies. Information. 2016;7:67.
39 Kazagli E, Chen J, Bierlaire M. Individual mobility analysis using smartphone data. Mobile Technologies for Activity-Travel Data Collection and Analysis2014. p. 187-208.
40 Wan N, Lin G. Life-space characterization from cellular telephone collected GPS data. Computers Environment and Urban Systems. 2013;39:63-70.
41 Brusilovskiy E, Klein LA, Salzer MS. Using global positioning systems to study health-related mobility and participation. Social Science & Medicine. 2016;161:134-42.
42 Glasgow ML, Rudra CB, Yoo EH, Demirbas M, Merriman J, Nayak P, et al. Using smartphones to collect time-activity data for long-term personal-level air pollution exposure assessment. J Expo Sci Environ Epidemiol. 2016;26:356-64.
43 Nguyen KT, Olgin JE, Pletcher MJ, Ng M, Kaye L, Moturu S, et al. Smartphone-Based Geofencing to Ascertain Hospitalizations. Circulation Cardiovascular quality and outcomes. 2017;10.
44 Palmer JR, Espenshade TJ, Bartumeus F, Chung CY, Ozgencil NE, Li K. New approaches to human mobility: using mobile phones for demographic research. Demography. 2013;50:1105-28.
45 Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res. 2013;15:e86.
46 Krenn PJ, Titze S, Oja P, Jones A, Ogilvie D. Use of global positioning systems to study physical activity and the environment: a systematic review. Am J Prev Med. 2011;41:508-15.
47 Bierlaire M, Chen JM, Newman J. A probabilistic map matching method for smartphone GPS data. Transportation Research Part C-Emerging Technologies. 2013;26:78-98.
48 Android Developers. Location | Android Developer. Available: https://developer.android.com/reference/kotlin/android/location/Location. Accessed: September 7, 2018.
49 Mazimpaka JD, Timpf S. Trajectory data mining: A review of methods and applications. Journal of Spatial Information Science. 2016;2016:61-99.
50 Xu R, Wunsch D, 2nd. Survey of clustering algorithms. IEEE Trans Neural Netw. 2005;16:645-78.

51 Palma AT, Bogorny V, Kuijpers B, Alvares LO. A clustering-based approach for discovering interesting places in trajectories. Proceedings of the 2008 ACM symposium on Applied computing - SAC '08; Fortaleza, Ceara, Brazil. 1363886: ACM; 2008. p. 863-8.
52 Yuan G, Sun PH, Zhao J, Li DX, Wang CW. A review of moving object trajectory clustering algorithms. Artificial Intelligence Review. 2017;47:123-44.
53 Zheng Y. Trajectory Data Mining: An Overview. Acm Transactions on Intelligent Systems and Technology. 2015;6.
54 Birant D, Kut A. ST-DBSCAN: An algorithm for clustering spatial–temporal data. Data & Knowledge Engineering. 2007;60:208-21.
55 Donaire-Gonzalez D, Valentin A, de Nazelle A, Ambros A, Carrasco-Turigas G, Seto E, et al. Benefits of Mobile Phone Technology for Personal Environmental Monitoring. JMIR Mhealth Uhealth. 2016;4:e126.
56 Sanchez M, Ambros A, Salmon M, Bhogadi S, Wilson RT, Kinra S, et al. Predictors of Daily Mobility of Adults in Peri-Urban South India. International journal of environmental research and public health. 2017;14:783.
57 Thierry B, Chaix B, Kestens Y. Detecting activity locations from raw GPS data: a novel kernel-based algorithm. Int J Health Geogr. 2013;12:14.
58 Liss DT, Serrano E, Wakeman J, Nowicki C, Buchanan DR, Cesan A, et al. "The Doctor Needs to Know": Acceptability of Smartphone Location Tracking for Care Coordination. JMIR Mhealth Uhealth. 2018;6:e112.
59 Carter ED, Ndhlovu M, Munos M, Nkhama E, Katz J, Eisele TP. Validity of maternal report of care-seeking for childhood illness. J Glob Health. 2018;8:010602.
60 Census of India. Administrative Atlas of India. New Delhi: Office of the Registrar General & Census Commissioner, 2011.
61 International Institute for Population Sciences (IIPS) and ICF. National Family Health Survey (NFHS-4), 2015-16: India. Mumbai: IIPS, 2017.
62 Mathew JL, Patwari AK, Gupta P, Shah D, Gera T, Gogia S, et al. Acute respiratory infection and pneumonia in India: a systematic review of literature for advocacy and action: UNICEF-PHFI series on newborn and child health, India. Indian pediatrics. 2011;48:191-218.
63 Shah D, Choudhury P, Gupta P, Mathew JL, Gera T, Gogia S, et al. Promoting appropriate management of diarrhea: a systematic review of literature for advocacy and action: UNICEF-PHFI series on newborn and child health, India. Indian pediatrics. 2012;49:627-49.
64 Pillai RK, Williams SV, Glick HA, Polsky D, Berlin JA, Lowe RA. Factors affecting decisions to seek treatment for sick children in Kerala, India. Soc Sci Med. 2003;57:783-90.
65 Awasthi S, Srivastava NM, Pant S. Symptom-specific care-seeking behavior for sick neonates among urban poor in Lucknow, Northern India. J Perinatol. 2008;28 Suppl 2:S69-75.
66 Mohan P, Iyengar SD, Agarwal K, Martines JC, Sen K. Care-seeking practices in rural Rajasthan: barriers and facilitating factors. J Perinatol. 2008;28 Suppl 2:S31-7.

67 Sreeramareddy CT, Sathyanarayana TN, Kumar HN. Utilization of health care services for childhood morbidity and associated factors in India: a national cross-sectional household survey. PLoS One. 2012;7:e51904.
68 Dongre AR, Deshmukh PR, Garg BS. Childhood morbidity, household practices and health care seeking for sick children in a tribal district of Maharashtra, India. Indian J Med Sci. 2010;64:7-16.
69 Ganatra B, Hirve S. Male bias in health care utilization for under-fives in a rural community in western India. Bull World Health Organ. 1994;72:101-4.
70 Google. Location Strategies | Android Developer. n.d. Available: https://developer.android.com/guide/topics/location/strategies.html. Accessed: June 11, 2017.
71 Zhou XL, Yu W, Sullivan WC. Making pervasive sensing possible: Effective travel mode sensing based on smartphones. Computers Environment and Urban Systems. 2016;58:52-9.
72 Montini L, Prost S, Schrammel J, Rieser-Schassler N, Axhausen KW. Comparison of travel diaries generated from smartphone data and dedicated GPS devices. Transport Survey Methods: Embracing Behavioural and Technological Changes. 2015;11:227-41.
73 Kestens Y, Thierry B, Chaix B. Re-creating daily mobility histories for health research from raw GPS tracks: Validation of a kernel-based algorithm using real-life data. Health Place. 2016;40:29-33.
74 International Institute for Population Sciences (IIPS) and Macro International. National Family Health Survey (NFHS-3), 2005-06: India: Volume 1. Mumbai: IIPS: 2007.
75 MEASURE DHS. Demographic and Health Surveys. Available: http://www.measuredhs.com/. Accessed: May 16, 2018.
76 International Institute for Population Sciences (IIPS). National Family Health Survey. Available: http://rchiips.org/nfhs/. Accessed: May 17, 2018.
77 UNICEF. Multiple Indicator Cluster Survey (MICS). Available: http://www.unicef.org/statistics/index_24302.html Accessed: May 16, 2018.
78 Roll Back Malaria. Malaria Indicator Surveys. Available: http://www.malariasurveys.org. Accessed: May 16, 2018.
79 Hazir T, Begum K, El Arifeen S, Khan AM, Huque MH, Kazmi N, et al. Measuring coverage in MNCH: a prospective validation study in Pakistan and Bangladesh on measuring correct treatment of childhood pneumonia. PLoS Med. 2013;10:e1001422.
80 Yoder PS, Hornik RC. Symptoms and perceived severity of illness as predictive of treatment for diarrhea in six Asian and African sites. Social Science & Medicine. 1996;43:429-39.
81 Malhotra N, Upadhyay RP. Why are there delays in seeking treatment for childhood diarrhoea in India? Acta Paediatr. 2013;102:e413-8.
82 Akin JS, Hutchinson P. Health-care facility choice and the phenomenon of bypassing. Health Policy Plan. 1999;14:135-51.
83 Ryan GW. What do sequential behavioral patterns suggest about the medical decision-making process?: modeling home case management of acute illnesses in a rural Cameroonian village. Soc Sci Med. 1998;46:209-25.

84 Nair H, Williams L, Marsh A, Lele P, Bhattacharjee T, Chavan U, et al. Assessing the impact of mobile phones and repeated surveys on healthcare seeking behaviour for common childhood illnesses in rural India. 2018.
85 Hirve S, Marsh A, Lele P, Chavan U, Bhattacharjee T, Nair H, et al. Concordance between GPS-based smartphone app for continuous location tracking and mother's recall of care-seeking for child illness in India. 2018.
86 Filmer D, Pritchett LH. Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India. Demography. 2001;38:115-32.
87 Rutstein SO, Johnson K. The DHS wealth index. Calverton, Maryland, USA: ORC Macro, 2004.
88 ESRI. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute; 2011.
89 Liang KY, Zeger SL. Regression analysis for correlated data. Annual review of public health. 1993;14:43-68.
90 StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP; 2015.
91 Smith F. The quality of private pharmacy services in low and middle-income countries: a systematic review. Pharm World Sci. 2009;31:351-61.
92 Chandy SJ. Consequences of irrational use of antibiotics. Indian journal of medical ethics. 2008;5:174-5.
93 Porter G, Grills N. Medication misuse in India: a major public health issue in India. Journal of public health (Oxford, England). 2016;38:e150-7.
94 Miller R, Goodman C. Performance of retail pharmacies in low- and middle-income Asian settings: a systematic review. Health Policy Plan. 2016;31:940-53.
95 Chowdhury F, Sturm-Ramirez K, Mamun AA, Iuliano AD, Bhuiyan MU, Chisti MJ, et al. Factors driving customers to seek health care from pharmacies for acute respiratory illness and treatment recommendations from drug sellers in Dhaka city, Bangladesh. Patient Prefer Adherence. 2017;11:479-86.
96 Manna B, Nasrin D, Kanungo S, Roy S, Ramamurthy T, Kotloff KL, et al. Determinants of health care seeking for diarrheal illness in young children in urban slums of Kolkata, India. Am J Trop Med Hyg. 2013;89:56-61.
97 Willis JR, Kumar V, Mohanty S, Singh P, Singh V, Baqui AH, et al. Gender differences in perception and care-seeking for illness of newborns in rural Uttar Pradesh, India. Journal of health, population, and nutrition. 2009;27:62-71.
98 Malhotra C, Malhotra R, Ostbye T, Subramanian SV. Maternal autonomy and child health care utilization in India: results from the National Family Health Survey. Asia Pac J Public Health. 2014;26:401-13.
99 Sikder SS, Labrique AB, Craig IM, Wakil MA, Shamim AA, Ali H, et al. Patterns and determinants of care seeking for obstetric complications in rural northwest Bangladesh: analysis from a prospective cohort study. BMC Health Serv Res. 2015;15:166.
100 Ganatra B, Hirve S. A community-based study on induced abortions. Issues Med Ethics. 2001;9:7-8.
101 Choi Y, El Arifeen S, Mannan I, Rahman SM, Bari S, Darmstadt GL, et al. Can mothers recognize neonatal illness correctly? Comparison of maternal report and assessment by community health workers in rural Bangladesh. Trop Med Int Health. 2010;15:743-53.

102 Marchant T, Bryce J, Victora C, Moran AC, Claeson M, Requejo J, et al. Improved measurement for mothers, newborns and children in the era of the Sustainable Development Goals. J Glob Health. 2016;6:010506.
103 Requejo JH, Newby H, Bryce J. Measuring coverage in MNCH: challenges and opportunities in the selection of coverage indicators for global monitoring. PLoS Med. 2013;10:e1001416.
104 Android Developers. Android Studio. Available: https://developer.android.com/studio/. Accessed: May 15, 2018.
105 Vich G, Marquet O, Miralles-Guasch C. Suburban commuting and activity spaces: using smartphone tracking data to understand the spatial extent of travel behaviour. Geographical Journal. 2017;183:426-39.
106 Shareck M, Kestens Y, Gauvin L. Examining the spatial congruence between data obtained with a novel activity location questionnaire, continuous GPS tracking, and prompted recall surveys. Int J Health Geogr. 2013;12:40.
107 Isaacson M, Shoval N, Wahl HW, Oswald F, Auslander G. Compliance and data quality in GPS-based studies. Transportation. 2016;43:25-36.
108 Kroeger A. Health interview surveys in developing countries: a review of the methods and results. Int J Epidemiol. 1983;12:465-81.
109 Vazquez-Prokopec GM, Bisanzio D, Stoddard ST, Paz-Soldan V, Morrison AC, Elder JP, et al. Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. PLoS One. 2013;8:e58802.
110 Searle KM, Lubinda J, Hamapumbu H, Shields TM, Curriero FC, Smith DL, et al. Characterizing and quantifying human movement patterns using GPS data loggers in an area approaching malaria elimination in rural southern Zambia. R Soc Open Sci. 2017;4:170046.
111 Qi F, Du F. Tracking and visualization of space-time activities for a micro-scale flu transmission study. Int J Health Geogr. 2013;12:6.
112 Beekhuizen J, Kromhout H, Huss A, Vermeulen R. Performance of GPS-devices for environmental exposure assessment. J Expo Sci Environ Epidemiol. 2013;23:498-505.
113 Xu Y, Shaw SL, Zhao ZL, Yin L, Lu F, Chen J, et al. Another Tale of Two Cities: Understanding Human Activity Space Using Actively Tracked Cellphone Location Data. Annals of the American Association of Geographers. 2016;106:489-502.
114 Curriero FC, Pinchoff J, van Landingham SW, Ferrucci L, Friedman DS, Ramulu PY. Alteration of travel patterns with vision loss from glaucoma and macular degeneration. JAMA Ophthalmol. 2013;131:1420-6.
115 Marsh A, Bhattacharjee T, Hirve S, Lele P, Chavan U, Bhujbal S, et al. TrackCare: A smartphone-based approach for tracking participant movement in rural Pune district, India. 2018.
116 Chen Y-C A Tutorial on Kernel Density Estimation and Recent Advances. ArXiv e-prints [Internet]. Available: https://ui.adsabs.harvard.edu/#abs/2017arXiv170403924C. 'Accessed.'
117 King TL, Bentley RJ, Thornton LE, Kavanagh AM. Using kernel density estimation to understand the influence of neighbourhood destinations on BMI. BMJ Open. 2016;6.

118 Stopher PR, Shen L, Liu W, Ahmed A. The Challenge of Obtaining Ground Truth for GPS Processing. Transport Survey Methods: Embracing Behavioural and Technological Changes. 2015;11:206-17.
119 Calu Costa J, Wehrmeister FC, Barros AJ, Victora CG. Gender bias in careseeking practices in 57 low- and middle-income countries. J Glob Health. 2017;7:010418.
120 Sengupta A, Nundy S. The private health sector in India. BMJ (Clinical research ed). 2005;331:1157-8.
121 Leena V Gangolli RD, Abhay Shukla. Review of Healthcare in India. Mumbai: Centre for Enquiry into Health and Allied Themes; 2005.
122 Math SB, Moirangthem S, Kumar CN. Public health perspectives in cross-system practice: past, present and future. Indian journal of medical ethics. 2015;12.

CURRICULUM VITAE

ANDREW DAVID MARSH

Born February 5, 1987 in Springfield, MA, USA

PERSONAL DATA

Department of International Health
615 N. Wolfe St.
Baltimore, MD 21205-2179
413.537.4178
amarsh6@jhmi.edu

EDUCATION AND TRAINING

Expected August 2018	Doctor of Philosophy, International Health Global Disease Epidemiology and Control Johns Hopkins School of Public Health Baltimore, MD Advisor: Professor Robert Black Thesis project: *Assessing the determinants of care-seeking for childhood illness in rural Pune district, Maharashtra state, India*
May 2012	Master of Public Health Johns Hopkins School of Public Health Baltimore, MD
May 2009	Bachelor of Arts, Economics The College of William and Mary Williamsburg, VA
January 2008 – June 2008	Study Abroad Foreign Student Study Center (CEPE) University of Guadalajara Guadalajara, Mexico

PROFESSIONAL EXPERIENCE

Position:	Graduate Research Assistant
Institution:	PI: Jennifer Bryce and Melinda Munos, Institute for International Programs Department of International Health Johns Hopkins Bloomberg School of Public Health Baltimore, MD
Dates:	May 2011 – Current
Principal Responsibilities:	• Supporting implementation of an mHealth study to validate maternal recall of care-seeking location in rural Pune district, India (April 2014 – present) • Collected and analyzed cost data for the Real-Time Mortality Monitoring (RMM) program (2011-2014), which piloted innovative methods of measuring under-five mortality in five African countries • Supported data collection of household- and facility-level costs associated with provision of child health services in Malawi (Summer 2013) • Supported RMM data management systems in Ghana (Summer 2012)
Position:	Economist
Institution:	Supervisor: Mark Zak, Survey of Occupational Injuries and Illnesses, Case and Demographic Characteristics Program Bureau of Labor Statistics, Department of Labor Washington, DC
Dates:	September 2009 – October 2011
Principal Responsibilities:	• Administered Survey of Occupational Injuries and Illnesses, which annually collects workplace injury and illness data from 200,000 private industry establishments • Co-authored annual publications highlighting case and demographic statistics for the most severe workplace injuries and illnesses
Position:	Intern Health Economist

Institution:	Supervisor: Shamim Qazi, Department of Maternal, Newborn, Child and Adolescent Health World Health Organization Geneva, Switzerland
Dates:	May 2009 - August 2009
Principal Responsibilities:	• Modeled the number of child deaths averted by implementing the Global Action Plan for the Prevention and Control of Pneumonia (GAPP) using the Lives Saved Tool • Calculated the cost of implementing the GAPP in 68 high-mortality countries
Position:	Youth Development Intern
Institution:	Supervisor: Elena Reilly Save the Children USA Managua, Nicaragua
Dates:	May 2009 - August 2009
Principal Responsibilities:	• Collaborated with project staff to improve an existing community health program • Supported community youth groups in development of microfinance-based business plans

AWARDS:

- Robert D. and Helen S. Wright Fellowship in International Health, 2013

REFEREE / REVIEW EXPERIENCE:

- *Journal of Global Health*

PUBLICATIONS:

1. Helleringer, S., Arhinful, D., Abuaku, B., Humes, M., Wilson, E., **Marsh, A.**, Clermont, A., Black, R. E., Amouzou, A. (2018). Using community-based reporting of vital events to monitor child mortality: Lessons from rural Ghana. *PLoS One, 13*(1), e0192034. doi:10.1371/journal.pone.0192034

2. Silva, R., Amouzou, A., Munos, M., **Marsh, A.**, Hazel, E., Victora, C., Black, R., Bryce, J. (2016). Can community health workers report accurately on births and deaths? Results of field assessments in Ethiopia, Malawi and Mali. *PLoS One, 11*(1). doi:10.1371/journal.pone

3. Hazel, E., Bryce, J., & the IIP-JHU iCCM Evaluation Working Group (including **Marsh, A.**) (2016). On Bathwater, Babies, and Designing Programs for Impact: Evaluations of the Integrated Community Case Management Strategy in Burkina Faso, Ethiopia, and Malawi. *Am J Trop Med Hyg, 94*(3), 568-570. doi:10.4269/ajtmh.94-3intro1

4. Bryce, J., Amouzou, A., Victora, C. G., Jones, G., Silva, R., Hill, K., Black, R. E., & the RMM Working Group (including **Marsh, A.)** (2016). "Real-Time" Monitoring of Under-Five Mortality: Lessons for Strengthened Vital Statistics Systems. *PLoS Med, 13*(1), e1001904. doi:10.1371/journal.pmed.1001904

5. Bryce, J., & the RMM Working Group (including **Marsh, A.**) (2016). "Real-Time" Monitoring of Under-Five Mortality: A Vision Tempered by Reality. *PLoS Med, 13*(1), e1001912. doi:10.1371/journal.pmed.1001912

6. Amouzou, A., Kanyuka, M., Hazel, E., Heidkamp, R., **Marsh, A.**, Mleme, T., Munthali, S., Park, L., Banda, B., Moulton, L., Black, R. E., Hill, K., Perin, J., Victora, C. G, Bryce, J. (2016). Independent Evaluation of the integrated Community Case Management of Childhood Illness Strategy in Malawi Using a National Evaluation Platform Design. *Am J Trop Med Hyg, 94*(3), 574-583. doi:10.4269/ajtmh.15-0584

7. Amouzou, A., Kanyuka, M., Hazel, E., Heidkamp, R., **Marsh, A.**, Mleme, T., Munthali, S., Park, L., Banda, B., Moulton, L., Black, R. E., Hill, K., Perin, J., Victora, C. G, Bryce, J. (2016). Independent Evaluation of the Integrated Community Case Management of Childhood Illness Strategy in Malawi Using a National Evaluation Platform Design. *Am J Trop Med Hyg, 94*(6), 1434-1435. doi:10.4269/ajtmh.16-0110b

8. **Marsh, A.**, Munos, M., Baya, B., Sanon, D., Gilroy, K., & Bryce, J. (2013). Using LiST to model potential reduction in under-five mortality in Burkina Faso. *BMC Public Health, 13*(Suppl.3), S26. doi:10.1186/1471-2458-13-S3-S26

MANUSCRIPTS UNDER REVIEW:

9. *Hirve, S.*, **Marsh, A.***, Lele, P., Chavan, U., Bhattacharjee, T., Nair, H., Campbell, H., Juvekar, S. Concordance between GPS-based smartphone app for continuous location tracking and mother's recall of care-seeking for child illness in India.*
   ** Joint first-author*

10. ***Marsh, A.**, Bhattacharjee, T., Hirve, S., Lele, P., Chavan, U., Bhujbal, S., Lin, H., Nair, H., Juvekar, S., Campbell, H. TrackCare: A smartphone-based approach for tracking participant movement in rural Pune district, India.*

11. *Nair, H., Williams, L., **Marsh, A.**, Lele, P., Bhattacharjee, T., Chavan, U., Hirve, S., Campbell, H., Juvekar, S. Assessing the impact of mobile phones and repeated surveys on healthcare seeking behaviour for common childhood illnesses in rural India.*

12. *Lele, P., Chavan, U., **Marsh, A.**, Hirve, S., Bhattacharjee, T., Campbell, H., Nair, H., Juvekar, S. Qualitative study of caregiver experience with use of a smartphone-based activity tracking application ("TrackCare") in a semi-urban community near Pune, India: impact on maternal recall of location of care-seeking.*

13. *Apte, A., Ingole, V., Lele, P., **Marsh, A.**, Bhattacharjee, T., Hirve, S., Campbell, H., Nair, H., Chan, S., Juvekar, S. Ethical Considerations in the Use of GPS-Based Movement Tracking in Health Research – Lessons from a Care-Seeking Study in Rural West India.*

TEACHING EXPERIENCE:

Johns Hopkins Bloomberg School of Public Health
Department of International Health

2013	Introduction to International Health Primary Instructor: Henry Perry Teaching assistant
2012	Case Studies in Primary Health Care Primary Instructor: Henry Perry Teaching assistant
2012	Large-Scale Effectiveness Evaluations of Health Programs Primary Instructor: Kate Gilroy Teaching assistant